

# Badania budżetów gospodarstw domowych w Polsce

Marek Pęczkowski, mgr, Wydział Nauk Ekonomicznych, Uniwersytet Warszawski

Badania budżetów gospodarstw domowych (BGD) stanowią podstawowe źródło informacji o poziomie i strukturze dochodów, wydatków, spożycia i wyposażenia w dobra trwałe gospodarstw domowych. Należą one do pierwszych badań warunków życia ludności w Polsce prowadzonych (od 1957 roku) metodą reprezentacyjną, pozwalającą na uogólnienie wyników na całą populację. Wyniki badania budżetów gospodarstw domowych są wykorzystywane do:

- analizy poziomu i zróżnicowania warunków życia podstawowych grup społeczno-ekonomicznych ludności,
- szczegółowych badań rynku konsumpcyjnego i spożycia naturalnego żywności,
- badania ubóstwa i wykluczenia społecznego,
- tworzenia wag do obliczeń indeksu cen towarów i usług konsumpcyjnych,
- podejmowania decyzji o charakterze społeczno-gospodarczym, np. ustalania najniższego wynagrodzenia i szacowania obciążeń podatkowych.

Na podstawie wyników tych badań przeprowadza się różnorodne analizy socjologiczne i ekonomiczne o charakterze naukowym oraz dokonuje się porównań międzynarodowych.

Badania budżetów polegają na prowadzeniu przez określony czas tzw. książeczek budżetowych, w których badane gospodarstwo rejestruje wszystkie przychody i rozchody pieniężne i niepieniężne wraz z określeniem ich rodzaju i źródeł, ilości zakupionych towarów oraz liczbę posiłków spożytych poza domem i liczbę posiłków spożytych w gospodarstwie przez osoby niebędące jego członkami. Oprócz tego przez ankietera na podstawie wywiadu osobistego zostaje wypełniana karta statystyczna gospodarstwa domowego dostarczająca informacji o cechach demograficznych i społeczno-ekonomicznych oraz aktywności zawodowej członków gospodarstwa, warunkach mieszkaniowych, wyposażeniu mieszkania w instalacje techniczno-sanitarne, powierzchni użytkowej gruntów rolnych, subiektywnej ocenie dochodu i warunków życia gospodarstwa. Przeprowadza się również wywiad końcowy w celu zebrania dodatkowych informacji o wyposażeniu gospodarstwa w dobra trwałe użytkowania oraz o zakupie niektórych dóbr (tzw. rozchody rzadkie). Dane uzyskane z kwestionariuszy i książeczek budżetowych są rejestrowane

na nośnikach komputerowych w wydziałach badań ankietowych urzędów statystycznych. Dane te są scalane i po przeprowadzeniu kontroli są wykorzystywane przez GUS do sporządzenia tablic prezentacyjnych, a także mogą być udostępniane jednostkom zewnętrznym w postaci plików komputerowych. Na podstawie tych danych GUS wydaje corocznie publikację *Budżety gospodarstw domowych*. Zawiera ona tablice dotyczące:

- przeciętnych miesięcznych przychodów i rozchodów (w rozbiciu na poszczególne pozycje) w zł na osobę,
- przeciętnego miesięcznego spożycia wybranych artykułów żywnościowych w jednostkach naturalnych (kg, l, szt.) na osobę,
- ilości wybranych dóbr trwałego użytkowania (w procentach) w gospodarstwach domowych.

Wartości w tablicach są podawane w różnych przekrojach, np. według liczby osób, typu gospodarstwa, grup dochodowych. W opracowaniach tych jest opisywana metodologia badań, zasady doboru gospodarstw do próby oraz najważniejsze definicje. Dokładny opis metodologii badań oraz definicji i klasyfikacji występujących kategorii znajduje się w opracowaniu [*Metodyka...*, 1998].

Wybrane wyniki badań budżetów są publikowane w wydawnictwach GUS, jak *Rocznik Statystyczny*, *Mały Rocznik Statystyczny*, *Rocznik Statystyczny Województw*, „Biuletyn Statystyczny” oraz w innych publikacjach GUS wydawanych w serii: „Informacje i opracowania statystyczne GUS” oraz „Studia i analizy statystyczne GUS”.

## 1. Historia badań

W latach 1957–1971 gospodarstwa domowe były wybierane do badań według specjalnego schematu losowania spośród pracowników zatrudnionych w wybranych zakładach pracy w gospodarce społecznej. Od 1973 roku zastosowano podejście terytorialne, polegające na doborze dwustopniowym. Najpierw utworzono w województwach tzw. jednostki losowania pierwszego stopnia (lps), osobno na terenach miejskich i wiejskich, i z nich losowano lps z prawdopodobieństwem proporcjonalnym do liczby mieszkań uzyskanych ze spisu 1970 roku. Potem w ramach wybranych lps losowano gospodarstwa domowe. Badano około 10 tys. gospodarstw wszystkich typów z wyjątkiem gospodarstw członków milicji, wojska oraz pracujących na rachunek własny.

Do 1982 roku badania BGD były prowadzone metodą ciągłą, tzn. to samo gospodarstwo domowe było badane przez cały rok lub dłużej. W latach 1983–1992 badania prowadzono metodą rotacji kwartalnej w cyklu rocznym. Od 1993 roku stosuje się metodę rotacji miesięcznej. Rotacja polega na tym, że jedno gospodarstwo jest badane przez jeden kwartał (obecnie miesiąc) w roku, a następnie część gospodarstw jest badana w analogicznym kwartale (miesiącu) w następnym roku. Początkowo przyjęto zasadę, że 50% wylosowanych gospodarstw uczestniczy w badaniach przez kolejne cztery lata (jest to tzw. próba stała), a 50% gospodarstw uczestniczy w badaniach tylko jedno-

krotnie (tzw. próba zmienna). Od 2001 roku gospodarstwa są badane tylko przez dwa kolejne lata, np. w 2007 roku badano dwie niezależne podpróbki wylosowane w 2005 roku:

- próbka 1. biorąca udział w badaniu 2006–2007,
- próbka 2. biorąca udział w badaniu 2007–2008.

Oznacza to, że w jednym roku jest wymieniana połowa badanych gospodarstw. Obecnie w jednym roku jest badanych ponad 30 000 gospodarstw domowych, co stanowi około 0,25% populacji.

Badanie budżetów gospodarstw domowych nie obejmuje tzw. gospodarstw zbiorowych, do których należą internaty, domy studenckie, domy opieki, klasztory. Do 1992 roku badaniami budżetów gospodarstw domowych obejmowano gospodarstwa domowe:

- pracownicze, których głównym źródłem utrzymania jest dochód z pracy najmniej w sektorze publicznym lub prywatnym,
- pracowniczochłopskie, tj. pracowników najemnych i użytkujących gospodarstwo rolne,
- chłopskie, utrzymujące się z pracy w indywidualnym gospodarstwie rolnym,
- emerytów i rencistów. Ta grupa obejmowała również gospodarstwa, których jedynym źródłem utrzymania jest inna forma pomocy społecznej niż emerytura i renta.

Nie badano gospodarstw, w których skład wchodziły osoby pracujące na własny rachunek (np. rzemieślników) lub wykonujące wolny zawód (np. prywatna praktyka lekarska lub adwokacka). Nie badano też gospodarstw obywateli państw obcych. Szacuje się, że badana populacja obejmowała około 85% ogółu ludności kraju.

Od 1993 roku objęto badaniami gospodarstwa pracujących na własny rachunek poza rolnictwem lub wykonujących wolny zawód. Spośród gospodarstw należących do grupy emerytów i rencistów wyodrębniono gospodarstwa utrzymujące się z niezarobkowych źródeł innych niż emerytura i renta (zasiłki dla bezrobotnych, zasiłki rodzinne, alimenty, dochody z własności).

Od 2005 roku gospodarstwa są klasyfikowane według pięciu podstawowych grup społeczno-ekonomicznych:

- gospodarstwa pracowników,
- gospodarstwa rolników,
- gospodarstwa pracujących na własny rachunek,
- gospodarstwa emerytów i rencistów,
- gospodarstwa utrzymujące się z niezarobkowych źródeł.

Ponadto wyróżnia się podgrupy: gospodarstwa pracowników zatrudnionych na stanowisku robotniczym i nierobotniczym oraz osobno gospodarstwa emerytów i gospodarstwa rencistów. Badane są gospodarstwa domowe obywateli obcych państw zamieszkujących w Polsce przez dłuższy czas i posługujące się językiem polskim, natomiast nie bada się gospodarstw członków korpusu dyplomatycznego państw obcych.

## 2. Metoda doboru próby

Stosowany jest terytorialny wielowarstwowy dwustopniowy schemat losowania. Jednostkami losowania pierwszego stopnia są terenowe punkty badań (tpb) utworzone na podstawie wykazu rejonów statystycznych, opracowanych dla potrzeb Narodowego Spisu Powszechnego (NSP). Rejony te są aktualizowane co roku o zmiany spowodowane nowym budownictwem, ubytkami w starym budownictwie i zmiany w podziale administracyjnym kraju. Małe rejony, które liczą mniej niż 250 mieszkań, są łączone w zespoły rejonów. W ten sposób zostają utworzone tpb. Jednostki te zostały podzielone na warstwy według województw, a w każdym województwie według klasy miejscowości. Do badania w 2007 roku utworzono łącznie 96 warstw, w tym 31 wiejskich, z których wylosowano 1566 tpb. Liczba tpb jest wybierana z każdej warstwy proporcjonalnie do liczby mieszkań w warstwie, co zapewnia, że każde mieszkanie ma jednakowe prawdopodobieństwo wyboru do próby.

Operat losowania w drugim stopniu stanowiły wykazy mieszkań w wylosowanych tpb. Wykazy te zostały opracowane przez wojewódzkie urzędy statystyczne. W każdym tpb losuje się jednakową liczbę mieszkań (24 mieszkania, po dwa na każdy miesiąc), ponadto losuje się próbę rezerwową liczącą do 150 mieszkań. W wylosowanych mieszkaniach są badane wszystkie gospodarstwa domowe.

## 3. Niepodejmowanie badań

Poważnym problemem w badaniach BGD jest niepodejmowanie badania przez gospodarstwa wylosowane do badań. W 2007 roku nie podjęło badania 44,7% gospodarstw domowych wylosowanych do badania po raz pierwszy lub gospodarstw badanych już poprzednim roku. Gospodarstwa niepodejmujące badania nie mają cech losowych, dlatego analizowane są przyczyny niepodejmowania badania przez gospodarstwa domowe. Przyczyny te można podzielić na przyczyny obiektywne (rozpad gospodarstwa, zmiana miejsca zamieszkania, utrudniony kontakt z członkami gospodarstwa ze względu na podeszły wiek lub chorobę, nieobecność w domu) lub przyczyny subiektywne (odmowa wzięcia udziału w badaniu ze względu na niechęć lub brak czasu do prowadzenia systematycznych zapisów przychodów i rozchodów). Przyczyny obiektywne najczęściej występują w gospodarstwach emerytów i rencistów, a przyczyny subiektywne w gospodarstwach pracowników i pracujących na własny rachunek, głównie wśród gospodarstw ludzi młodych i o małej liczbie osób. Udział gospodarstw przerywających prowadzenie zapisów lub nieudzielających wywiadu kwartalnego nie jest duży i wynosi około 1%.

Gospodarstwa nieprzystępujące do badania są zastępowane przez gospodarstwa z próby rezerwowej, jednak nie muszą być to gospodarstwa o podobnych cechach społeczno-demograficznych (np. mogą mieć inne główne źródło utrzymania lub inną liczbę osób). W związku z tym zostały wyznaczone wagi mające na celu dostosowanie struktury próby ważonej do struktury popula-

cji. W 2007 roku przy konstrukcji wag brano pod uwagę wielkość gospodarstwa (liczbę osób) oraz podział na miasto i wieś zgodnie z danymi pochodzącymi z NSP 2002 roku. Wszelkie analizy powinny być przeprowadzane na próbie ważonej.

#### **4. Zawartość danych**

Dane składają się z dwóch rodzajów:

- dane o gospodarstwach domowych, w których jednostką (obserwacją) jest gospodarstwo domowe,
- dane o osobach, w których jednostką jest osoba będąca członkiem gospodarstwa.

Wśród członków gospodarstwa wyróżnia się głowę gospodarstwa — osobę odniesienia, która osiąga najwyższy dochód spośród wszystkich członków gospodarstwa.

Dane o gospodarstwach zawierają:

- informacje ogólne i socjologiczne: miesiąc badania, województwo, klasę miejscowości, powierzchnię mieszkania, powierzchnię gospodarstwa rolnego, informacje o mieszkaniu (np. formę własności), główne i dodatkowe źródło utrzymania, typ rodziny (np. gospodarstwo jednoosobowe, małżeństwo bez dzieci, małżeństwo z dziećmi, samotna matka lub ojciec z dziećmi);
- informacje o wyposażeniu mieszkania w podstawowe urządzenia, jak kuchnia, łazienka, ciepła woda, telefon stacjonarny, garaż;
- informacje o wyposażeniu gospodarstwa w dobra trwałe, np. telewizor, komputer, samochód (obecnie jest rejestrowanych 39 pozycji dóbr trwałych);
- przychody netto. Są to wszystkie wartości wpływające do gospodarstwa domowego bez uwzględnienia zaliczek na podatek dochodowy od osób fizycznych płaconych przez płatnika w imieniu podatnika oraz bez składek na ubezpieczenie społeczne i zdrowotne. Obejmują dochody otrzymywane z różnych tytułów (z pracy najemnej, z pracy na własny rachunek, z indywidualnego gospodarstwa rolnego, ze świadczeń z ubezpieczeń społecznych i pomocy społecznej, z wynajmu nieruchomości, prywatnych darów i alimentów). Pozostałe pozycje przychodowe obejmują m.in. pobrane lokaty oszczędnościowe, zacięgnięte pożyczki i kredyty, sprzedaż majątku rzeczowego. Uwzględnia się przychody pieniężne i niepieniężne, np. wartości artykułów pobranych z gospodarstwa rolnego na własne potrzeby, wartości artykułów otrzymanych bezpłatnie lub częściowo odpłatnie;
- rozchody netto. Są to wszystkie wartości wpływające z gospodarstwa domowego na zewnątrz. Składają się z wydatków konsumpcyjnych, wydatków pozostałych i pozycji rozchodowych (wpłacone lokaty, spłacone kredyty, składki ubezpieczeniowe na życie, pożyczki udzielone osobom prywatnym, zakup lub modernizacja majątku trwałego, podatki);



- spożycie ilościowe wybranych produktów, głównie żywności i alkoholu. Dane o spożyciu nie obejmują artykułów spożywczych konsumowanych w placówkach gastronomicznych, dlatego rejestrowane spożycie wielu artykułów (w tym alkoholu) jest zaniżone w stosunku do danych makroekonomicznych.

Dane o osobach zawierają:

- informacje demograficzne: płeć, wiek, stan cywilny, stopień pokrewieństwa z głową gospodarstwa,
- informacje o poziomie wykształcenia i nauce szkolnej,
- informacje o aktywności zawodowej: zawód, forma zatrudnienia, status bezrobocia,
- indywidualne dochody przypisane do osób: wynagrodzenia, emerytura i renta, zasiłek dla bezrobotnych, inne zasiłki.

## 5. Panel

Na początku lat dziewięćdziesiątych ubiegłego stulecia podjęto próby doprowadzenia do porównywalności badań polskich budżetów gospodarstw domowych z danymi międzynarodowymi. W ramach projektu PACO (*Panel Comparability*) na Uniwersytecie Warszawskim przygotowano czteroletni panel gospodarstw badanych w latach 1987–1990 ([Górecki i in., 1996]). Wszedł on w skład zintegrowanej bazy zawierającej dane o dochodach gospodarstw domowych i dane społeczno-demograficzne wielu krajów. Struktura danych zawartych w ujednoliconej bazie PACO jest identyczna dla każdego kraju uczestniczącego w projekcie, jednakowe są również definicje poszczególnych zmiennych. Polskie dane kwartalne zostały przekształcone do danych rocznych metodą imputacji po przeliczeniu ich według indeksów kosztów utrzymania z poszczególnych miesięcy i skorygowaniu ze względu na sezonowość. Na podstawie istniejących kategorii dochodów i innych cech gospodarstwa zostały utworzone kategorie dochodowe porównywalne z kategoriami dochodowymi w innych krajach. W ostatecznej wersji baza PACO zawiera dane z wybranych lat z zakresu 1983–1997, dotyczące 8 krajów. Są to: Francja, Niemcy, Węgry, Luksemburg, Polska, Hiszpania, Wielka Brytania, USA. Polskę reprezentują dwa panele: dla lat 1987–1990 (3707 gospodarstw) i dla lat 1994–1996 (4958 gospodarstw).

Następnie w ramach projektu CHER (*Consortium of the Household Panels for European Socio-economic Research*) utworzono dwa panele: 1994–1996 i 1997–2000 [Cieciela, Pęczkowski, 2001]. Pierwszy z tych paneli jest trzyletni, gdyż zdecydowano się opuścić rok 1993 ze względu na odmienną klasyfikację dochodów oraz niską jakość danych, związaną ze zmianą metodologii badań budżetów w Polsce. Celem projektu CHER, podobnie jak PACO, jest stworzenie międzynarodowej, porównywalnej bazy danych panelowych. Projekt CHER jest w pewnym stopniu kontynuacją projektu PACO, jednak zasadniczo została zmieniona organizacja plików oraz lista i definicje zmiennych. Podstawowe zmiany to wprowadzenie znacznie szerszej informacji o aktyw-

ności zawodowej i zatrudnieniu, warunkach mieszkaniowych, wydatkach i relacjach społecznych. Ostateczna baza CHER zawiera dane z 17 krajów (Austria, Belgia, Dania, Finlandia, Francja, Niemcy, Grecja, Węgry, Irlandia, Włochy, Luksemburg, Holandia, Polska, Portugalia, Hiszpania, Szwajcaria, Wielka Brytania) obejmująca lata 1990–2000. Zmienne dostępne w bazie CHER to: zmienne organizacyjne (16), wagi (4), zmienne demograficzne (14), aktywność zawodowa (13), zatrudnienie (24), edukacja (6), zdrowie (7), mieszkanie (19), wyposażenie w dobra trwałe (7), dochody (30), wydatki (5), subiektywna ocena warunków życia (5), relacje społeczne (4).

Polskie panele były wykorzystywane przez badaczy do międzynarodowych porównań rozkładu i struktury dochodów, analizy mobilności dochodów i badania zmiany dochodów ze względu na zmiany w strukturze społeczno-demograficznej gospodarstw (np. przybycia nowej osoby, urodzenia się dziecka, ubycia osoby, utraty pracy, wzrostu poziomu wykształcenia).

Od 2001 roku nie jest możliwe utworzenie paneli dłuższych niż dwuletnie ze względu na zmiany zasad doboru gospodarstw domowych do próby. Nie można przez to przeprowadzić wielu analiz przekrojowo-czasowych.

## 6. Uwagi o porównaniu z danymi makroekonomicznymi

Od 1993 roku oprócz zmiany metody losowania próby GUS zmieniał klasyfikację przychodów i rozchodów oraz klasyfikację typów społeczno-ekonomicznych dostosowując ją do standardów międzynarodowych. Wprowadzono pojęcie dochodu rozporządzalnego. Jest to suma bieżących dochodów gospodarstwa pomniejszona o zaliczki na podatek dochodowy od osób fizycznych płacone przez płatnika w imieniu podatnika, o podatki od dochodów i własności płacone przez osoby pracujące na własny rachunek i użytkujące gospodarstwo rolne oraz o składki na ubezpieczenie społeczne i zdrowotne.

Na podstawie próby gospodarstw domowych prowadzi się wiele badań modułowych. Polegają one na tym, że w wybranych gospodarstwach (na ogół badanych w ustalonym miesiącu) są zbierane dodatkowe dane tematyczne. W latach 1993–2006 przeprowadzono następujące badania modułowe [Kordos, 2008]: pomoc społeczna (1993), potrzeby żywieniowe (1993), warunki mieszkaniowe (1993), potrzeby edukacyjne (1994), gospodarstwo domowe w gospodarce rynkowej (1994), zdrowie rodziny i wydatki na ochronę zdrowia (1994), wyposażenie gospodarstw domowych w dobra trwałego użytkowania (1995), zaspokojenie potrzeb rodziny (1995), uczestnictwo w turystyce i kulturze (1995), korzystanie z wideo (1998), ochrona zdrowia (1998), turystyka i wypoczynek (1998, 2001, 2005), ubezpieczenia osobowe i majątkowe w gospodarstwach domowych (1998, 2004), skutki powodzi w gospodarstwach domowych (1998), uczestnictwo w sporcie i rekreacji ruchowej (1999), potrzeba oszczędzania w gospodarstwach domowych (1999), ochrona zdrowia w gospodarstwach domowych (1999, 2003), spożycie żywności w gospodarstwach domowych i badania antropometryczne (2000), badanie zużycia paliw i energii w gospodarstwach domowych (2002), uczestnictwo w kulturze (2004), ścieżki

edukacyjne Polaków (2004), wykorzystanie technologii informacyjno-telekomunikacyjnej w gospodarstwach domowych (2005), ochrona zdrowia w gospodarstwach domowych (2006).

Istnieje rozbieżność pomiędzy wynikami badań budżetów gospodarstw domowych a danymi makroekonomicznymi. Jest to spowodowane odmienną metodologią badań, odmiennymi definicjami wielu kategorii (np. dochodu, liczby osób w gospodarstwie), odmiennymi klasyfikacjami (np. pozycje żywnościowe w badaniach budżetów nie obejmują żywienia w placówkach gastronomicznych, szpitalach, szkołach i przedszkolach, zatem są one zaniżone w stosunku do danych makroekonomicznych). Ponadto obserwuje się, że w badaniach reprezentacyjnych opartych na oświadczeniach osób badanych zaniżone są niektóre pozycje, jak dochody o charakterze niestałym, wydatki na alkohol i tytoń, żywienie w placówkach gastronomicznych. Z tego względu nie można bezpośrednio porównywać wyników badań budżetów gospodarstw domowych z danymi makroekonomicznymi. Dane pochodzące z budżetów gospodarstw domowych nadają się w szczególności do analiz zróżnicowania względnego dochodów, wydatków i spożycia oraz wyposażenia w dobra trwałego użytkowania ze względu na różne cechy społeczno-demograficzne (liczba osób, źródło utrzymania, klasa miejscowości, typ rodziny, wiek i poziom wykształcenia głowy gospodarstwa itp.) oraz zróżnicowania terytorialnego (województwa i regiony).

## 7. Przykład analizy

Jako przykład analizy opartej na danych z budżetów gospodarstw w 2007 roku przedstawimy badanie struktury wydatków konsumpcyjnych gospodarstw w zależności od różnych cech gospodarstwa. Wydatki na towary i usługi konsumpcyjne zostały podzielone na 12 grup:

- 1) żywność i napoje bezalkoholowe,
- 2) napoje alkoholowe, wyroby tytoniowe i narkotyki,
- 3) odzież i obuwie,
- 4) użytkowanie mieszkania i nośniki energii,
- 5) wyposażenie mieszkania,
- 6) zdrowie,
- 7) transport,
- 8) łączność,
- 9) rekreacja i kultura,
- 10) edukacja,
- 11) restauracje i hotele,
- 12) pozostałe towary i usługi konsumpcyjne.

Klasyfikacja ta jest zgodna z międzynarodową klasyfikacją COICOP/HBS (*Classification of Individual Consumption by Purpose for Household Budget Survey*). Szczegółowa zawartość poszczególnych kategorii jest podana w opracowaniu [Budżety..., 2008].



Utworzone zostały zmienne oznaczające udziały poszczególnych kategorii wydatków w wydatkach konsumpcyjnych ogółem. Są to zmienne: *zynw, alko, odzi, mies, wypo, zdro, tran, lacz, kult, eduk, rest, inne*.

Grupy decylowe dochodu na osobę zostały utworzone osobno w każdym miesiącu, co jest związane z prezentowaniem danych o dochodach w ujęciu nominalnym i występującą inflacją oraz sezonowością. Następnie grupy decylowe dla poszczególnych miesięcy zostały połączone w jedną grupę decylową.

Tabela 1. przedstawia strukturę 12 grup wydatków konsumpcyjnych według grup decylowych dochodu rozporządzalnego na osobę. Ostatni wiersz tabeli 1. zawiera strukturę wydatków konsumpcyjnych wszystkich gospodarstw domowych.

## Tabela 1.

Profile wydatków konsumpcyjnych gospodarstw domowych w 2007 według grup decylowych dochodu rozporządzalnego na osobę

Grupa decylowa dochodu na osobę	ZYWN	ALKO	ODZI	MIES	WYPO	ZDRO	TRAN	LACZ	KULT	EDUK	REST	INNE
1	0,448	0,037	0,041	0,188	0,042	0,031	0,052	0,048	0,044	0,006	0,017	0,046
2	0,404	0,036	0,046	0,201	0,042	0,037	0,058	0,053	0,052	0,007	0,013	0,050
3	0,379	0,033	0,046	0,210	0,042	0,044	0,063	0,055	0,055	0,007	0,012	0,054
4	0,360	0,032	0,046	0,215	0,045	0,051	0,060	0,056	0,058	0,009	0,013	0,055
5	0,347	0,031	0,047	0,219	0,045	0,056	0,063	0,057	0,057	0,009	0,014	0,055
6	0,333	0,029	0,046	0,224	0,043	0,065	0,062	0,056	0,060	0,010	0,015	0,057
7	0,323	0,028	0,045	0,225	0,045	0,069	0,062	0,058	0,065	0,009	0,015	0,057
8	0,304	0,028	0,047	0,225	0,049	0,071	0,065	0,057	0,068	0,009	0,018	0,060
9	0,279	0,028	0,052	0,215	0,050	0,068	0,076	0,059	0,077	0,009	0,021	0,065
10	0,225	0,029	0,066	0,187	0,059	0,060	0,098	0,059	0,099	0,013	0,031	0,073
Ogółem	0,325	0,030	0,049	0,212	0,047	0,058	0,068	0,056	0,067	0,009	0,018	0,059

Patrząc na tabelę 1., możemy zauważyć, że wraz ze wzrostem numeru grupy decylowej dochodu spada udział wydatków na żywność, spada również udział wydatków na alkohol i tytoń, natomiast udział wydatków na kulturę i edukację generalnie wzrasta. W pierwszych i ostatnich grupach decylowych zaburzana jest monotoniczność niektórych pozycji (np. mieszkanie, zdrowie).

Współzależność struktury wydatków i cech społeczno-demograficznych można przedstawić w postaci tabeli, ale dla dużej liczby kategorii staje się ona mało czytelna. Wygodniej jest przedstawić wyniki w formie graficznej w postaci tzw. mapy percepcji. Do utworzenia takich map wykorzystamy ideę zbliżoną do metody analizy korespondencji. Analiza korespondencji w kla-

sycznym ujęciu pozwala odwzorować związki między kategoriami dwóch zmiennych jakościowych w postaci układu punktów w przestrzeni wielowymiarowej, a następnie rzutować je w przestrzeń o mniejszej liczbie wymiarów (najczęściej na płaszczyznę) z jak najmniejszą stratą informacji o konfiguracji punktów. Wówczas rozważane zależności można przedstawić w formie graficznej umieszczając na jednym wykresie punkty odpowiadające kategoriom obu zmiennych, co ułatwia interpretację tych zależności [Greenacre, 1993; Stanimir, 2005].

Punktem wyjścia klasycznej analizy korespondencji jest tablica kontyngencji dwóch zmiennych jakościowych. Analiza korespondencji może być zastosowana do badania współzależności układu zmiennych stanowiących strukturę określonej cechy ciągłej i obiektów lub grup obiektów [Hair i in., 1998]. Panek [2008] nazywa takie podejście nieklasyczną analizą korespondencji. U nas grupy obiektów to grupy gospodarstw wyodrębnionych ze względu na określoną cechę jakościową.

W analizie korespondencji wyjściową tablicę (jak tab. 1.) przekształca się w tablicę korespondencji  $P = (p_{ij})$ , powstałą przez podzielenie każdego elementu przez sumę wartości wszystkich elementów. Na podstawie macierzy  $P$

wyznacza się macierz profili wierszowych  $R = (r_{ij}) = \frac{p_{ij}}{\sum_j p_{ij}}$  (są to rozkłady

warunkowe cechy wierszowej; w naszym przypadku profile pokrywają się z udziałami poszczególnych wydatków w sumie wydatków) oraz macierz profili kolumnowych  $C = (c_{ij}) = \frac{p_{ij}}{\sum_i p_{ij}}$ . Profile wierszowe mogą być odwzorowane

jako układ  $r$  punktów w przestrzeni  $s$ -wymiarowej (gdzie  $r$  — liczba grup obiektów,  $s$  — liczba zmiennych, u nas  $s = 12$  pozycji wydatków). Profile kolumnowe mogą być odwzorowane jako układ  $s$  punktów w przestrzeni  $r$ -wymiarowej. Ponadto wyznacza się średni profil wierszowy i średni profil kolumnowy, odpowiadające rozkładowi brzegowym odpowiednio wierszy i kolumn. Punkty reprezentowane przez średnie profile są nazywane centroidami i umieszcza się je w środku układu współrzędnych odpowiednio przestrzeni wierszowej i kolumnowej.

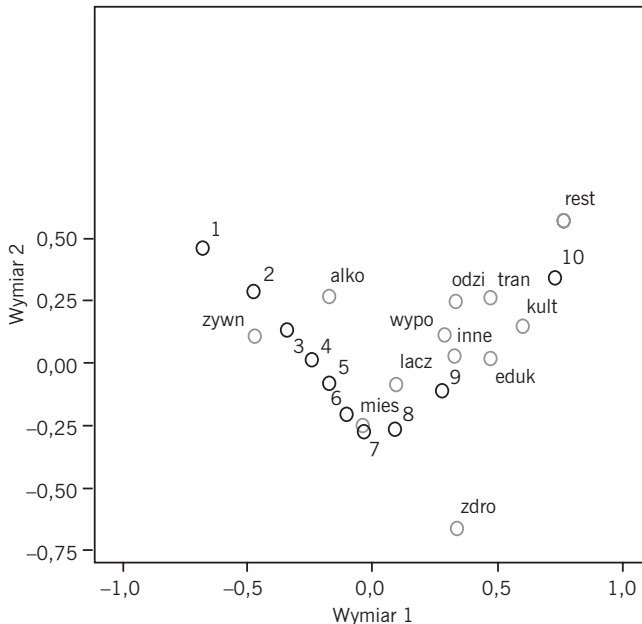
W przestrzeni wierszowej i kolumnowej definiuje się odległość między punktami za pomocą metryki  $\chi^2$ . Jest to ważona metryka euklidesowa, w której wagi są zależne od częstości brzegowych. Użycie metryki  $\chi^2$  zamiast metryki euklidesowej zapobiega zbyt dużemu wpływowi najliczniejszych kategorii na odległości. Z odległościami  $\chi^2$  związane jest pojęcie bezwładności. Bezwładność jest miarą rozproszenia punktów w stosunku do środka układu współrzędnych. Można ją rozłożyć na sumę bezwładności cząstkowych względem poszczególnych wymiarów. Następnie znajdujemy najlepszą płaszczyznę rzutowania w ten sposób, aby odpowiadające jej osie (wymiar 1, wymiar 2) wyjaśniały jak największą część całkowitej bezwładności. Udział bezwład-

ności cząstkowych w całkowitej bezwładności jest miarą jakości odwzorowania konfiguracji punktów na płaszczyźnie.

Interpretacja punktów na wykresie jest następująca:

- jeżeli punkt jest położony bardzo blisko początku układu współrzędnych, to znaczy że wyznaczony dla niego profil niewiele różni się od profilu średniego,
- punkty położone daleko od początku układu współrzędnych opisują profile różniące się znacznie od profilu średniego,
- bliskie położenie dwóch punktów odpowiadającym zmiennym lub grupom obiektów wskazuje, że ich profile są bardzo podobne,
- analiza położenia punktów odpowiadającym zmiennym lub grupom obiektów względem początku układu współrzędnych wskazuje na stopień zróżnicowania zmiennych lub grup,
- dalekie położenie punktu dla zmiennych od początku układu współrzędnych świadczy o silnym zróżnicowaniu odpowiadającej mu zmiennej,
- dalekie położenie punktu dla grupy od pozostałych punktów dla grup wskazuje na nietypowość danej grupy ze względu na strukturę opisaną przez zmienne,

W analizie korespondencji nie można bezpośrednio interpretować odległości między punktami reprezentującymi zmienne i grup obiektów, ponieważ są to przedstawione na jednym wykresie punkty należące do dwóch różnych przestrzeni rzutowania. Można jednak interpretować położenie punktów gru-



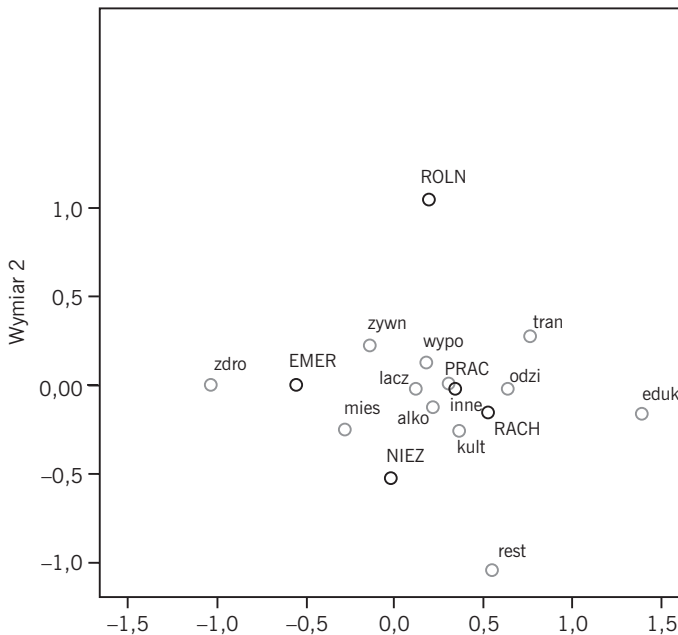
**Rys. 1.**

Mapa percepcji struktury wydatków konsumpcyjnych według grup decylowych dochodu na osobę

py względem konfiguracji punktów zmiennych lub położenie punktów zmiennych względem konfiguracji punktów grupy. Bliskie położenie punktu odpowiadającego zmiennej i punktu odpowiadającego grupie oznacza, że istnieje silne powiązanie między daną zmienną i grupą.

Rys. 1. przedstawia mapę percepcji dla danych przedstawionych w tabeli 1. Liczby na wykresie oznaczają numery grup decylowych.

Na rys. 2. i 3. zostały przedstawione mapy percepcji struktury wydatków konsumpcyjnych względem typu gospodarstwa i według poziomu wykształcenia głowy gospodarstwa. Podobne mapy można utworzyć wybierając inne grupy gospodarstw, np. według wielkości gospodarstwa, klasy miejscowości, województwa zamieszkania, typu rodziny, wieku głowy gospodarstwa, stanu cywilnego głowy gospodarstwa.

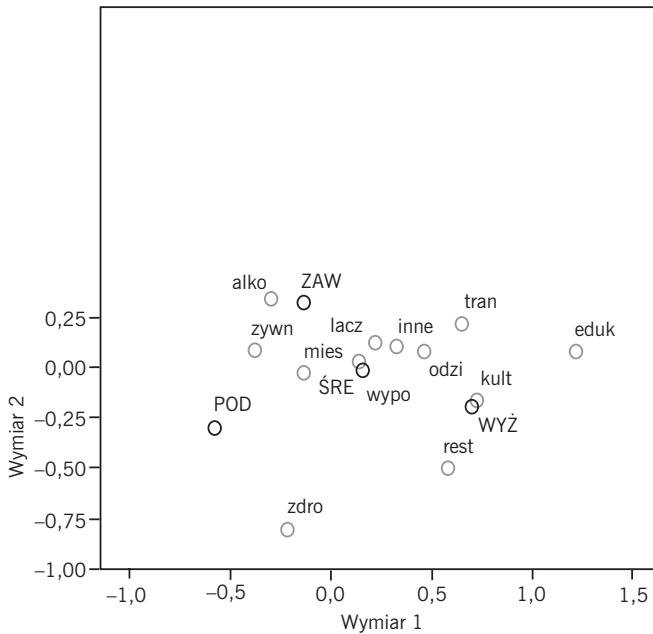


**Rys. 2.**

Mapa percepcji struktury wydatków konsumpcyjnych według typu gospodarstwa (pracownicy, rolnicy, pracujący na własny rachunek, emeryci i renciści, utrzymujący się ze źródeł niezarobkowych)

Gospodarstwa emerytów i rencistów mają ze wszystkich grup największy udział wydatków na zdrowie (9,3%), a najmniejszy udział wydatków na edukację (0,2%). Gospodarstwa osób utrzymujących się z niezarobkowych źródeł mają największy udział wydatków na alkohol i tytoń (4,1%), restauracje i bary (3,8%), mieszkanie (25,0%). Struktura wydatków gospodarstw rolników najbardziej odbiega od profilu średniego. Jest to związane z metodologią badania budżetów, według której rejestrowana jest wartość produktów pobranych

z gospodarstwa rolnego. Gospodarstwa rolników zdecydowanie mniej niż inne gospodarstwa wydają na żywienie w restauracjach, barach i stołówkach.



**Rys. 3.**

Mapa percepcji struktury wydatków konsumpcyjnych według poziomu wykształcenia głowy gospodarstwa: wyższe, średnie, zawodowe, podstawowe)

Gospodarstwa osób z wyższym wykształceniem można scharakteryzować przez duży udział wydatków na kulturę, restauracje, edukację i transport. Można zauważyć podobieństwo współzależności struktury wydatków z wykształceniem wyższym i z najwyższą grupą decylową dochodu z tym, że tutaj większe znaczenie mają wydatki na edukację i kulturę. Udział wydatków na zdrowie jest nietypowy. Te wydatki są najbardziej związane z wykształceniem podstawowym (*POD*). Powodem jest współzależność wydatków na zdrowie i wieku głowy gospodarstwa oraz typu gospodarstwa (emeryci i renciści). Na uwagę zwraca bliskie położenie punktów *alko* i *ZAW*.

Rzutując punkty na oś pierwszego wymiaru możemy nadać zmiennym lub grupom charakter porządkowy. Widzimy, że w przypadku grup decylowych wynik uzyskany w analizie korespondencji pokrywa się z rzeczywistym uporządkowaniem grup decylowych. Podobnie jest w przypadku poziomu wykształcenia. Typ gospodarstwa (rys. 2.) jest cechą jakościową, natomiast po rzutowaniu na oś 1 mamy kolejność: *EMER*, *NIEZ*, *ROLN*, *PRAC*, *RACH*, co można uznać za uporządkowanie typów gospodarstwa ze względu na strukturę wydatków konsumpcyjnych.



## Bibliografia

- Budżety gospodarstw domowych w 2007 r.*, (2008), „Informacje i opracowania statystyczne”, GUS, Warszawa.
- Cieciela J., Pęczkowski M., 2001, *Bazy danych dostępne na WNE UW. Wstępna charakterystyka*, dostępne na stronie <http://coin.wne.uw.edu.pl/panel/panel.html>.
- Górecki B., Pęczkowski M., Grodner A., 1996, *Polish Household Panel 1987–1990 as PACO Dataset*, „Research Paper” No.19, Luxembourg.
- Greenacre M., 1993, *Correspondence Analysis in Practice*, London, Academic Press.
- Hair J.F., Anderson R.E., Tatham R.L., Black W.C., 1998, *Multivariate Data Analysis*, Prentice Hall.
- Kordos J., 1997, *40 lat badań budżetów gospodarstw domowych w Polsce*, „Wiadomości Statystyczne” nr 7.
- Kordos J., 2008, *Rozwój badań warunków życia ludności*, „Wiadomości Statystyczne” nr 2.
- Metodyka badania budżetów gospodarstw domowych*, 1999, „Zeszyty Metodyczne i Kłasyfikacyjne”, GUS, Warszawa.
- Panek T., 2008, *Ubóstwo i wykluczenie społeczne w Polsce w układzie wojewódzkim*, w: *Statystyka wczoraj, dziś i jutro*, „Biblioteka Wiadomości Statystycznych” t. 56, GUS PTS, Warszawa.
- Stanimir A., 2005, *Analiza korespondencji jako narzędzie do badania zjawisk ekonomicznych*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.

## **A b s t r a c t** The Household Budget Survey in Poland



Conducted by the Central Statistical Office (CSO) household budget survey (HBS) is a major source of revenue and expenditure, and consumption of durable goods equipment of Polish households. These data is also essential for many advanced statistical analysis and international comparisons. The article presents the objectives and scope of HBS with a brief review of history and methodology. Particular attention was paid to the contribution of other staff of the Faculty of Economic Sciences participating in international projects PACO and CHER using the panel data based on the annual collection of the CSO. Other studies conducted in Poland using the HBS data were mentioned as well. At the end an example of multivariate statistical analysis based on data from the HBS 2007 was presented. This is the study of the consumption expenditure of households using the correspondence analysis.