

Henryk Domański*
Instytut Filozofii i Socjologii PAN

WIELONOMIALNY MODEL LOGITOWY. PRZYKŁAD ZASTOSOWANIA I INTERPRETACJA

Przystępując do analiz zależności między zmiennymi trzeba dokonać wyboru techniki statystycznej. Musi być ona dostosowana do typu skali, którą zmienne reprezentują, od tego bowiem zależy trafność sformułowanych ustaleń. Najważniejszy jest z tego punktu widzenia podział na zmienne ilościowe i jakościowe, zwane nominalnymi: decyzja o wprowadzeniu do analizy tych ostatnich zawęży pole wyboru, ponieważ zmienne nominalne reprezentują najniższy poziom pomiaru i w odniesieniu do nich niektórymi technikami statystycznymi nie można się posługiwać. Modele logitowe, o których tu będzie mowa, znajdują zastosowanie w sytuacji, gdy zmienna nominalna jest „wyjaśnianym” zjawiskiem.

Analiza logitowa jest uogólnieniem problemu regresji; chodzi w niej o ustalenie siły i kierunku zależności pomiędzy zmienną Y a predyktorami X_i . Modele logitowe stosowane są najczęściej wtedy, gdy Y jest zmienną dychotomiczną, np. gdy chodzi o pokazanie, w zależności od czego ludzie zaliczają się np. do „klasy robotniczej”, a nie do „klasy średniej”, albo w analizach nad preferencjami wyborczymi, aby „wyjaśnić”, kto deklaruje, że będzie głosował, powiedzmy, na „liberatów”. W obydwu przypadkach zmienna zależna operacjonalizowana jest w postaci zero-jedynkowej. Respondenci, którzy zaliczają się do klasy średniej otrzymują 1, a wszyscy pozostali 0; analogicznie można też zdefiniować zamiar głosowania w wyborach, przypisując 1 osobom stwierdzającym, że będą głosować, dajmy na to, na Unię Wolności.

Co robić, kiedy racje teoretyczne przemawiają za ujmowaniem wyjaśnianego zjawiska w postaci zestawu jakościowo odmiennych kategorii, których nie da się sprowadzić do dychotomicznego podziału? Chcemy np. zbadać, od czego zależy zróżnicowanie preferencji wyborczych ujmowanych w podziale na liberatów, socjaldemokratów, ludowców, populistów i przedstawicieli innych opcji. Przedstawiam poniżej jedno z rozwiązań tego problemu, a mianowicie wielonomialną analizę logitową. Model wielonomialny częściej jest wykorzystywany w ekonometrii niż w socjologii, gdzie zaczęto go stosować później i jest on stosunkowo mniej znany. Pokażę na konkretnych przykładach, na jakie pytania odpowiada wielonomialna analiza logitowa w porównaniu z modelem dychotomicznym oraz jakie są jej możliwości i ograniczenia interpretacyjne.

* Uwagi do autora lub prośby o nadbitki prosimy kierować do: Henryk Domański, Instytut Filozofii i Socjologii PAN, 00-330 Warszawa, ul. Nowy Świat 72, tel. (22) 65 72 297, E-mail: hdomansk@ifispan.waw.pl.

RODZAJ PROBLEMU

Zagadnieniom związanym z uogólnieniem analizy logitowej na model wielonomialny (ang. *multinomial logit regression*) poświęcono kilka prac z zakresu ekonometriki (zob. Theil 1969; Cragg i Uhler 1970; Schmidt i Strauss 1975; Madalla 1983; Green 1992). Modeli takich jest kilka. Najbardziej znany nosi nazwę wielonomialnego w ogólnej postaci; na nim się tutaj skoncentruję.

Założmy, że interesuje nas rozkład preferencji wyborczych w Polsce. Pytanie, na które poszukujemy odpowiedzi, dotyczy tego, kto głosuje na określone partie polityczne: kogo częściej wybierają ludzie młodzi, jak głosują mężczyźni w porównaniu z kobietami, mieszkańcy wielkich miast w przeciwieństwie do mieszkańców wsi, jak silnie zależy to od ich statusu materialnego, religijności i od poziomu wykształcenia. W odniesieniu do ostatnich wyborów do parlamentu we wrześniu 1997 r. mogłaby to być np. analiza wpływu tych charakterystyk na zamiar głosowania na 6 partii: AWS, SLD, PSL, Unię Wolności, na Krajową Partię Emerytów i Rentistów i na Unię Pracy.

Do tak postawionego problemu można podejść z dwojakiej perspektywy, stosując dwa różne schematy analiz. Pierwszy z nich, wykorzystywany znacznie częściej, sprowadza się do przeprowadzenia sześciu analiz logitowych. W każdym z sześciu modeli regresji wyjaśnianym zjawiskiem jest deklaracja o głosowaniu na AWS, SLD itd., ujmowana w kategoriach „tak-nie”. Estymowane są zależności pomiędzy tymi deklaracjami a płcią, wiekiem, wykształceniem i pozostałymi charakterystykami wprowadzonymi do modeli. Ich wpływ wyrażony jest w wartościach współczynników regresji — uzyskane w ten sposób rezultaty pozwalają wnioskować o zróżnicowaniu preferencji wyborczych elektoratu. Takim schematem postępowali się autorzy znanych studiów nad preferencjami wyborczymi. Przykładem jego zastosowań mogą być analizy przedstawione w książce *Understanding Political Change. The British Voter, 1967–1987*, której autorzy starają się rozstrzygnąć głośną w socjologii ostatnich lat kontrowersję wokół problemu: czy na przestrzeni ostatnich dziesięcioleci wpływ przynależności klasowej na preferencje wyborcze zmniejsza się, czy też utrzymuje się na zbliżonym poziomie, bez większych zmian (Heath i in. 1991).

Dychotomiczna regresja logitowa pozwala odstąpić mechanizmy zależności — ale dla każdej partii z osobna. Nie interesuje nas w tym przypadku wpływ zróżnicowania społecznego na preferencje polityczne jednostek ujmowane w postaci pełnego przekroju postaw: w ogóle nie brana jest tu pod uwagę możliwość, że ludzie postrzegają scenę polityczną jako rozkład wielu wyborów, nie zaś w postaci kontrastów typu „tak-nie”. Jeśli celem analizy ma być ustalenie czynników rządzących wyborem dokonywanym spośród wielu opcji, a nie tylko wyboru jednej z dwóch alternatyw, to można na to pytanie odpowiedzieć stosując drugi schemat, oparty na wielonomialnej analizie logitowej. Zestaw dychotomicznych zmiennych wyjaśnianych zastępujemy w tym przypadku jedną zmienną, złożoną z więcej niż dwóch wartości. W rozpatrywanym przykładzie byłby to rozkład preferencji wyborczych, zdefiniowanych jako zmienna nominalna, złożona z sześciu kategorii, poczynając od głosowania na AWS, kończąc na UP. W modelu wielonomialnym chodzi o „wyjaśnienie” kształtu tego rozkładu: na przykład, z jaką siłą

i w jakim kierunku różnicują go takie cechy jak płeć, wiek, poziom wykształcenia czy dochody.

Jakie informacje pozwalają na formułowanie tego rodzaju wniosków? Przypomnijmy najpierw najogólniejszą postać modelu logitowego. Charakteryzuje on stopień dopasowania krzywej — będącej funkcją wartości zmiennej Y względem wektora predyktorów X — do obserwowanego rozkładu. Matematyczną postać krzywej logitowej wyraża formuła:

$$\log(P/1-P) = a + b_1X_1 + b_2X_2 + \dots + b_kX_k, \quad (1)$$

gdzie $\log(P/1-P)$, jest dychotomiczną zmienną wyjaśnianą, określoną przez logarytm stosunku prawdopodobieństw — występowania i niewystępowania analizowanego zjawiska — zaś wartości b_1, \dots, b_k są współczynnikami regresji dla predyktorów X_1, \dots, X_k . Wyjaśnienie kluczowego dla regresji logitowej problemu — dlaczego zmienna Y zostaje zastąpiona ilorazem prawdopodobieństw — przekracza zadania niniejszego tekstu. Można je znaleźć w zamieszczonej na końcu literaturze przedmiotu.

Funkcja logitowa jest jedną z możliwych postaci krzywej nazywanej sigmoidalną, ponieważ swym kształtem przypomina ona grecką literę sigma. Krzywa ta stosunkowo najlepiej aproksymuje rozkład zależności w sytuacjach, gdy zmienna wyjaśniana jest dychotomią, w porównaniu z funkcją liniową, której najlepszym odzwierciedleniem jest klasyczny model regresji. Jeżeli chodzi o porównanie z innymi krzywymi sigmoidalnymi, to funkcja logistyczna jest stosunkowo najłatwiejsza do zastosowania w operacjach matematycznych i pozwala na oszacowanie parametrów modelu łatwych w interpretacji — z tych powodów model logitowy wykorzystywany jest w analizach najczęściej. Szczegółowe omówienie jego formalnych właściwości można znaleźć w pracach Maddali (1983) i Demarisa (1992), przy czym należy jeszcze raz powtórzyć, że istnieją alternatywne postaci krzywej sigmoidalnej; jedną z nich jest np. kumulatywny rozkład normalny, będący podstawą probitowego modelu regresji (zob. Aldrich i Nelson 1984).

Modele logitowe w wersji dychotomicznej i wielonomialnej dostarczają dwóch podstawowych statystyk. Po pierwsze, współczynników regresji, które analogicznie jak w regresji liniowej, mogą być interpretowane jako wskaźniki wpływu zmiennych X na wyjaśniane zjawisko Y — o ile oczywiście istnieją teoretyczne przesłanki do takiej interpretacji. Drugą podstawową informacją jest prawdopodobieństwo określające stopień dopasowania rozpatrywanego modelu do faktycznych zależności — przy założeniu, że jest to prawdziwy model. Wielkość tej statystyki oznaczanej dużą literą L (od ang. *likelihood*) zawiera się w przedziale od 0 do 1; większa wartość L świadczy o lepszym dopasowaniu. Jak zobaczymy za chwilę na przykładzie danych z Polski, standardowy pakiet analiz komputerowych dla modeli dychotomicznych podaje przekształconą wartość L , wygodniejszą w interpretacji. Natomiast w modelu multinomialnym, L prezentowane jest w logarytmicznej postaci (logarytmu naturalnego, przy podstawie e), ponieważ logarytmami łatwiej jest się posługiwać. Logarytm L jest liczbą ujemną, jako że L przyjmuje wartości mniejsze od 1.

Likelihood jest wynikiem empirycznego testu. W analizie wielonomialnej testowana jest hipoteza, czy badane zjawisko faktycznie stanowi układ kilku kategorii

zróznicowanych pod względem predyktorów X_i, \dots, X_n . Formułując problem od strony merytorycznej można powiedzieć, że dotyczy on strukturalizacji zjawisk, jeśli rozumieć pod tym terminem wyodrębnianie się kategorii społecznych, określonych przez różnice płci, wieku, wykształcenia i innych atrybutów. Wielkość *likelihood* dostarcza kryterium rozstrzygającego o znaczeniu strukturalizacji, na „tak” lub „nie”. Jeśli L przewyższa przyjęty poziom istotności statystycznej, to hipotezę zerową, która głosi, że strukturalizacja nie występuje, trzeba odrzucić. Natomiast otrzymanie wyniku z odpowiednio niskim *likelihood* nie daje podstaw, aby sądzić, że w analizowanej sferze zjawisk wyodrębnia się układ kategorii definiowanych w terminach zmiennych X .

Głównym celem wielonomialnej analizy logitowej jest zatem weryfikacja sformułowanego przez badacza modelu zależności, który ma „wyjaśnić” (w statystycznym sensie) występowanie rozpatrywanego zjawiska w postaci więcej niż dwóch jakościowo odmiennych kategorii. Jeśli test wypada pozytywnie, można przejść do interpretacji współczynników regresji, które informują, jakie charakterystyki jednostek tkwią u podstaw wyjaśnianego rozkładu.

Modele logitowe estymowane są przy pomocy metody *maximum-likelihood* — nie można się tu posłużyć, jak w regresji liniowej, metodą najmniejszych kwadratów. Punktem wyjścia jest sformułowanie funkcji określającej prawdopodobieństwo (L) otrzymania wyniku z analizowanej próby w sytuacji, gdy prawdziwy jest zakładany model zależności. Zakłada się, że model w postaci wyrażonej przez równanie 1 jest poprawny, jednak nie znane są wartości parametrów a i b_i . Chodzi o otrzymanie a i b_i , które maksymalizują L — parametry te określa się mianem najlepiej dopasowanych. Metoda *maximum-likelihood* sprawdza pod tym kątem wszystkie możliwe kombinacje a i b_i , estymuje wartość L dla każdej z nich i dokonuje wyboru kombinacji zapewniającej najwyższą wartość L . Procedurę tę omawia Eliason (1993) charakteryzując szczegółowo właściwości metody *maximum-likelihood*.

Podsumujmy. Wielonomialna analiza logitowa wychodzi poza model, w którym wyjaśniane zjawisko sprowadzone jest do dychotomicznego podziału. Jak zawsze chodzi o to, aby posługiwać się trafnymi wskaźnikami analizowanych zjawisk. W sytuacji, gdy klasyczny model logitowy wymusza ich zdefiniowanie w dychotomicznej postaci, model wielonomialny pozwala uniknąć tego rodzaju uproszczeń.

PRZYKŁAD ZASTOSOWANIA

Dane

Praktyczne zastosowanie wielonomialnej analizy logitowej przedstawimy na przykładzie preferencji wyborczych. Materiałem ilustracyjnym będą dane z badania ankietowego przeprowadzonego przez Centrum Badania Opinii Społecznej w maju 1997 r., na 5 miesięcy przed wyborami do Sejmu i Senatu. Był to jeden z serii sondaży przedwyborczych realizowanych przez CBOS. W kwestionariuszu znalazło się jeszcze kilka innych bloków tematycznych, a wśród nich powtarzany stale blok pytań o nastroje, opinie na temat rządu i o stosunek do zmian w Polsce po 1989 r. Badanie przeprowadzono na ogólnokrajowej, losowo dobranej repre-

zencacji ludności mieszkańców Polski w wieku powyżej 18 lat. Z założonej próby 1200 osób (jednostką losowania były adresy rodzin) zrealizowano ostatecznie 1142 wywiadów, uwzględniając w tej liczbie jednostki z próby rezerwowej.

Blok pytań o preferencje wyborcze znalazł się już na drugiej stronie kwestionariusza. Respondentom zadano najpierw pytanie czy wzięliby udział w wyborach do Sejmu i Senatu, gdyby odbywały się one w najbliższą niedzielę. Po uzyskaniu odpowiedzi twierdzącej ankietę wręczał badanym osobom kartę z nazwami partii i pytał: „Na kandydatów, której partii lub ugrupowania głosował(a)by Pan(i) w wyborach?”. Respondenci wskazywali jedną z dziesięciu możliwości: (1) AWS, (2) BBWR (Bezpartyjny Blok dla Polski), (3) KPEiR (Krajowe Porozumienie Emerytów i Rencistów), (4) PSL, (5) ROP (Ruch Odbudowy Polski), (6) SLD, (7) Unię Polityki Realnej, (8) Unię Pracy, (9) Unię Wolności i (10) inne partie lub organizacje.

Jeśli chce się ustalić, od czego zależały deklarowane przez respondentów wybory, analiza logitowa wydaje się być właściwą metodą. Wyboru partii nie da się rozpatrywać inaczej niż jako zmiennej złożonej z jakościowo odmiennych kategorii. Nie układają się one w żadną hierarchię, o której można by powiedzieć, że wyraża natężenie preferencji wyborczych w określonym wymiarze. Jest to zatem typowa zmienna nominalna; posłużenie się w tym przypadku regresją liniową nie byłoby wskazane.

W naszym przykładzie nie obejmujemy analizą wszystkich dziesięciu, a tylko 6 kategorii, wyłączając 4 najrzadziej reprezentowane, których ze względu na niewielką liczebność nie można było rozpatrywać. Weźmiemy pod uwagę wskazania na: AWS łącznie z BBWR (identyfikujące sympatie „postsolidarnościowe” o nachyleniu centroprawicowym), SLD, PSL, Unię Wolności, KPEiR i Unię Pracy. Respondenci wybierający pozostałe partie zostali wyłączeni z analiz.

Co do wyznaczników zamiarów wyborczych, to z obiektywnych cech położenia społecznego zostały uwzględnione płeć (mężczyznom przypisywano wartość 1, kobietom 2), wykształcenie (definiowane przez liczbę ukończonych klas szkolnych: od 4 do 16 klas), miesięczne dochody na osobę w rodzinie (w tys. zł.) oraz wiek w podziale na kategorie: ludzi „młodych” (do 30 lat), „średnich” (31–50) i „starych” (powyżej 50 roku życia); ostatnia z wymienionych kategorii jest tzw. grupą odniesienia, pomijaną w modelach regresji — „odniesione” są do niej wartości dwóch pozostałych zmiennych zero-jedynkowych definiujących wiek.

Z rzeczy ważnych, które mogą wywierać pewien wpływ na deklarowany wybór partii warto uwzględnić przekonania i poglądy zwolenników różnych opcji. W Polsce ważna jest religijność; respondenci odpowiadali na pytanie, jak często biorą udział w praktykach religijnych, takich jak np. msze — ich odpowiedzi, od „tak, kilka razy w tygodniu” do „w ogóle w nich nie uczestniczę”, tworzą skalę religijności, której przypisano wartości od 1 do 5. Drugim rodzajem postaw rzutujących na decyzje wyborcze będą samooceny statusu ekonomicznego badanych osób: ich wskaźnikiem była odpowiedź na pytanie, „jak P. ocenia obecnie warunki materialne swojego gospodarstwa domowego?”. Skala ocen zawierała się w granicach od 1 („złe”) do 5 („dobre”).

Wziąłem też pod uwagę stosunek do socjalizmu, syntetyczny konstrukt utworzony na podstawie analizy czynnikowej z trzech pytań: (i) czy P. zdaniem od roku 1989, w Polsce, dużo zmieniło się („na lepsze, trochę na lepsze, trochę na gorsze, dużo na gorsze, praktycznie nic się nie zmieniło”), (ii) czy zmiany zachodzące

w Polsce od roku 1989 przynoszą ludziom więcej strat czy korzyści („same korzyści, ..., same straty”), (iii), czy z perspektywy czasu, warto czy też nie warto było zmieniać ustrój w Polsce (skala czterech odpowiedzi od „zdecydowanie warto” do „zdecydowanie nie warto”). Okazuje się, że opinie respondentów w tych kwestiach tworzą jeden wymiar, od zdecydowanego poparcia dla przemian (równoznacznego z krytyką socjalizmu) do ich negacji. Wymiar ten potraktowałem jako zmienną, wprowadzając ją do analiz.

O preferencjach wyborczych może też decydować stosunek do polityki ekonomicznej rządu i wizja pożądanego wizerunku sceny publicznej, znajdująca odzwierciedlenie w odpowiedziach na często powtarzane w badaniach ankietowych pytania w rodzaju: czy państwo powinno zapewnić obywatelom wysoki poziom świadczeń społecznych, czy powinno się ograniczać zarobki osób najwyżej zarabiających, czy należy się otworzyć na wymianę z zagranicą, czy Kościół powinien być oddzielony od państwa itp. W kwestionariuszu znalazła się bateria 10 pytań precyzujących poglądy w tych kwestiach, które kontrastują zwolenników rozwiązań prozachodnich i prorynkowych z ich przeciwnikami. Narzędziem specyfikacji różnic w tym wymiarze będzie zmienna „wizja zmian systemowych”, otrzymana po zastosowaniu analizy czynnikowej do tych pytań — skalę tę utworzono z wag „najsilniejszego” czynnika.

MODEL DYCHOTOMICZNY

Analityczne możliwości wielonomialnej analizy logitowej przedstawimy przez porównanie z wynikami analiz otrzymanych w modelu dychotomicznym. Będzie on punktem wyjścia.

Tabela 1.
Zależności pomiędzy deklaracjami głosowania na AWS, SLD, PSL, UW, KPEiR i UP w wyborach do parlamentu w 1997 r. a wybranymi cechami respondentów. Współczynniki regresji logitowej w modelu dychotomicznym

	Deklaracja głosowania (1=tak; 0=nie) na:					
	AWS	SLD	PSL	UW	KPEiR	UP
Płeć (1=mężczyźni; 2=kobiety)	- 0,39*	- 0,10	- 0,25	- 0,33	0,12	0,81*
Ludzie młodzi	- 0,18	- 0,80	- 0,73	0,45	- 8,11**	0,62
Ludzie w średnim wieku	- 0,03	- 0,38	0,13	0,07	- 1,26**	0,93
Liczba lat nauki	0,08	0,00	0,08	0,31**	- 0,03	0,13
Dochody rodzin	0,00	0,00	- 0,003*	0,00	0,00	0,00
Ocena pozycji materialnej	- 0,02	0,32	0,32**	0,04	- 0,15	0,05
Chodzenie do kościoła	- 0,76**	0,46**	- 0,22	- 0,01	0,33*	0,33*
Stosunek do socjalizmu	- 0,64**	0,15	- 0,03	- 0,54**	- 0,04	- 0,22
Wizja zmian systemowych	- 0,37**	0,13	0,02	0,35*	- 0,03	- 0,04

Wyraz stały	Deklaracja głosowania (1=tak; 0=nie) na:					
	0,24	-3,42	0,82	-6,14	-1,15	7,59
Chi-kwadrat/G ²	117/737	64/734	31/335	50/299	48/302	31/335

** p<0,01, * p<0,05

Liczba stopni swobody dla wszystkich równań = 9

Wyniki regresji wielorakiej dla 6 partii zdefiniowanych w postaci dychotomicznych wyborów przedstawiono są w tabeli 1. Współczynniki regresji logitowej mają identyczną interpretację jak w regresji liniowej, np. ujemna wartość dla płci w deklaracjach dotyczących wyboru AWS wskazuje, że kobiety rzadziej sympatyzowały z AWS niż mężczyźni. Różnica ta wyrażona jest w postaci logarytmu pomiędzy stosunkami prawdopodobieństw: tzn. pomiędzy P — wybraniem AWS — i (1-P), tj. niewybraniem tej partii. Różnica wynosi -0,39 przy kontroli wpływu wieku, wykształcenia i wszystkich innych zmiennych, które w tym modelu zostały uwzględnione. Jeżeli chodzi o wpływ wykształcenia, to awans o każdy kolejny poziom zwiększa logarytm stosunku tych prawdopodobieństw o 0,08. Religijność, przejawiająca się w chodzeniu do kościoła, związana jest negatywnie z P/(1-P) — osoby rzadziej praktykujące rzadziej zapowiadały głosowanie na AWS (-0,76). Natomiast wpływ przynależności do kategorii osób najmłodszych i w średnim wieku wyraża się odpowiednio wartościami -0,18 i -0,03 — można powiedzieć, że o tyle rzadziej deklarowali oni zamiar głosowania na AWS w porównaniu z najstarszymi respondentami, potraktowanymi w tym modelu jako kategoria odniesienia. Odejmując drugą z tych wielkości od pierwszej otrzymujemy „efekt” różnic w natężeniu sympatii wyborczych do AWS pomiędzy „młodymi” i „średnimi”. Przy „wyrównaniu” wpływu innych czynników wynosi ona 0,15; AWS miał najmniejsze szanse poparcia przez ludzi młodych.

Interpretując ściśle wielkości parametrów obrazujących związki poszczególnych zmiennych z wyborem AWS należałoby posługiwać się terminem „logit”, tak bowiem definiowana jest wielkość $\log P/(1-P)$, występująca po lewej stronie równania regresji (1). „Efekty” płci, religijności, wykształcenia i kategorii wieku, są logitami o wielkościach: -0,39, 0,76, 0,08, -0,18, -0,03 i 0,15, dla P/(1-P) w regresji preferencji wyborczych AWS. Problem polega na tym, że pojęcie logitu nie funkcjonuje w słownictwie socjologicznym na co dzień i nie ma tak bezpośredniego przełożenia na język zależności, jak np. określenie „przyrost (lub spadek wartości) zmiennej Y na jednostkę X_i”, którego używa się w regresji liniowej. Bardziej dostępne intuicyjnemu rozumieniu jest ujmowanie „efektów” jako wielkości prawdopodobieństwa „występuje/niewystępuje”, P/(1-P), przypadającej na jednostkę zmiennej niezależnej, bez odwoływania się do pojęcia logarytmów. Taka interpretacja współczynników regresji ma miejsce po sprowadzeniu funkcji logistycznej (będącej addytywną funkcją Y względem X_i, jak w równaniu (1)), do postaci multiplikatywnej:

$$P/(1-P) = e^a + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (2)$$

gdzie e jest podstawą logarytmu naturalnego, równą 2,71828.... Przekształcenie to jest konsekwencją faktu, że funkcja logarymiczna jest odwrotnością funkcji wy-

kładniczej, a logarytm b_k równy jest wykładnikowi potęgi b_k przy podstawie e (zob. Retherford i Choe 1993: 123–125). Parametry równania regresji w formie multiplikatywnej mają, w matematycznym sensie, znaczenie wielokrotności. Na przykład e^{bX} dla efektu płci wskazuje, o ile razy fakt bycia kobietą zmniejsza iloraz „wybrania/niewybrania” AWS w porównaniu z mężczyznami (jeśli wyeliminować oddziaływanie pozostałych zmiennych w modelu). Są to tzw. *odds ratio* (nazwa wzięta z literatury anglojęzycznej, co odpowiada mniej więcej polskiemu „stosunek przewag”) — w rozpatrywanym modelu dla AWS wielkość *odds* dla płci wynosi 0,68. W pakietach komputerowych SPSS i SAS, wartości *odds* zamieszczane są w wynikach obliczeń niezależnie od wartości logitów (w SPSS figurują one w ostatniej kolumnie zestawienia wyników).

Za wyborem multiplikatywnej wersji modelu przemawiają względy interpretacyjne, ale z innego punktu widzenia jest on trudniejszy do interpretacji. Należy pamiętać, że funkcja multiplikatywna nie jest funkcją liniową, co powoduje, że siła „efektu” poszczególnych zmiennych nie jest wartością stałą; zmienia się zależnie od wartości X_i ; z tego względu, przy omawianiu zależności wygodniejszy (bo uogólnia siłę związku przy pomocy jednego parametru) jest model w addytywnej postaci. Przy prezentacji ustaleń można się odwoływać zarówno do współczynników regresji z modelu addytywnego, jak i do *odds*, jako że dostarczają one ekwiwalentnych informacji co do istoty analizowanych zjawisk.

Czego dowiadujemy się z wyników analizy przedstawionej w tabeli 1?

Przed wszystkim tego, że rozpatrywane cechy położenia społecznego i postawy, wszystkie razem, pozostawały w znaczącym związku z prawdopodobieństwem oddania głosu na każdą z sześciu partii w nadchodzących wyborach. Syntetycznym wskaźnikiem występowania globalnego efektu są statystyki chi-kwadrat i *goodness of fit* (G^2), odniesione do liczby stopni swobody. Obydwie one mówią o tym samym, o stopniu dopasowania analizowanego modelu (zakładanych w nim zależności) do faktycznych uwarunkowań preferencji wyborczych, odtworzonych na podstawie danych z próby. Wielkości tych statystyk przewyższają krytyczną wartość chi-kwadrat, która dla dziewięciu stopni swobody wynosi 27,9 przy poziomie istotności $p = 0,001$. Znajdują się one w obszarze krytycznym, prawdopodobieństwo ich otrzymania jest minimalne, jeżeli hipoteza o braku zależności jest prawdziwa. Można ją zatem odrzucić.

O ile rozpatrywane zjawiska, razem wzięte, różnicują postawy wyborcze elektoratu, to część z nich, potraktowanych osobno, nie odgrywała istotnej roli. Na przykład płęć była tylko znaczącym czynnikiem wyboru AWS i UP, oddziałując zresztą na odwrót: UP miała wyraźnie sfeminizowany elektorat, o czym informuje dodatkna wartość współczynnika regresji, podczas gdy głosy na AWS znacząco częściej oddawali mężczyźni. Za deklarowanymi preferencjami dla różnych partii stoją więc nieco inne wzory zależności, co dowodzi odmienności ich elektoratu. Są to najogólniejsze wnioski, które można wysnuć z porównania wyników 6 dychotomicznych modeli regresji logitowej.

MODEL WIELONOMIALNY; INTERPRETACJA WSPÓŁCZYNNIKÓW REGRESJI

Przejsie do modelu wielonomialnego pociąga za sobą zasadniczą zmianę w interpretacji parametrów regresji. Sześć zmiennych wyjaśnianych, będących przedmiotem naszego zainteresowania, redukuje się teraz do jednej. Jest nią rozkład preferencji wyborczych dla 6 partii, z których jedna musi być potraktowana jako kategoria odniesienia. Nie istnieje bowiem żadna inna możliwość pokazania siły, z jaką zmienne wyjaśniające różnicują wartości zmiennej nominalnej — jeżeli chcemy ją wyrazić w terminach parametrów regresji. W analizie regresji zmienne muszą być „mieralne” na skalach co najmniej interwałowych. Można powiedzieć, że z tego względu wyodrębnienie kategorii odniesienia dla zmiennej nominalnej „wzmacnia” jej poziom pomiaru, pozwala bowiem określić porządek i natężenie wartości.

W dychotomicznym modelu logitowym kategorią odniesienia dla wyjaśnianych preferencji był fakt ich odrzucenia: wszyscy respondenci, którzy nie wybrali AWS lub SLD, bądź jednej z kolejnych czterech partii, otrzymywali 0. Przyjmijmy, przechodząc do modelu wielonomialnego, że kategorią odniesienia będzie elektorat UP — do niego zostaną odniesione wielkości parametrów ustalonych dla pozostałych 5 alternatyw. Tak jak dotychczas, badamy zależność preferencji partyjnych od dziewięciu zmiennych.

Wielonomialny model dla preferencji wyborczych zdefiniowany jest przez 5 następujących równań:

$$\log (P_1/P_6)=a_1 + b_1S + c_1W + \dots + m_1Z, \quad (3)$$

$$\log (P_2/P_6)=a_2 + b_2S + c_2W + \dots + m_2Z, \quad (4)$$

$$\log (P_3/P_6)=a_3 + b_3S + c_3W + \dots + m_3Z, \quad (5)$$

$$\log (P_4/P_6)=a_4 + b_4S + c_4W + \dots + m_4Z, \quad (6)$$

$$\log (P_5/P_6)=a_5 + b_5S + c_5W + \dots + m_5Z, \quad (7)$$

i przez równanie określające warunek,

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 = 1 \quad (8)$$

gdzie P_1, \dots, P_5 odnoszą się odpowiednio do prawdopodobieństw głosowania na AWS, SLD, PSL, UW i KPEIR, P_6 — do pominiętego w modelu wyboru Unii Pracy, a_i jest wyrazem stałym, b_iS identyfikuje zależność preferencji wyborczych od płci, c_iW — od wykształcenia, a m_iZ — od wizji zmian systemowych badanych osób; ze względu na brak miejsca, w równaniach (3) — (7) nie przedstawiono wszystkich 9 zmiennych wyjaśnianych, które zostały uwzględnione w modelu.

Tabela 2.

Zależności pomiędzy deklaracjami głosowania na AWS, SLD, PSL, UW i KPEiR w wyborach do parlamentu w 1997 r. a wybranymi cechami respondentów. Współczynniki regresji logitowej w modelu wielonomialnym

	Deklaracja głosowania na:				
	AWS	SLD	PSL	UW	KPEiR
Płeć (1=mężczyźni; 2=kobiety)	0,00	0,00	-0,01**	0,00	0,00
Ludzie młodzi	0,04	0,26	0,07	0,06	-0,16
Ludzie w średnim wieku	-0,94**	0,00	-0,41**	-0,22	-0,51**
Liczba lat nauki	0,00	0,00	0,00	0,01	0,00
Dochody rodzin	0,00	0,00	-0,003**	0,00	0,00
Ocena pozycji materialnej	-1,08*	-1,60**	-1,18**	-0,25	-0,15
Chodzenie do kościoła	-1,12	-1,29**	-0,92	-0,95	-2,30**
Stosunek do socjalizmu	0,00	0,00	0,00	0,00	0,00
Wizja zmian systemowych	-0,11	-0,16*	-0,22*	0,14	-0,25**
Wyraz stały	5,86	3,58	6,23	-0,48	6,40

** $p < 0,01$, * $p < 0,05$

Chi-kwadrat=269; Log L=-799; Liczba stopni swobody = 45

Parametry modelu wielonomialnego zamieszczone są w tabeli 2. Wielkości chi-kwadrat i log-likelihood, równe 269 i -799, przy 45 stopniach swobody, świadczą o dopasowaniu modelu do danych: można sprawdzić, że wartość krytyczna już dla 20 stopni swobody wynosi 50,9 przy bardzo wysokim poziomie istotności 0,01. Oznacza to, że zestaw rozpatrywanych zmiennych różnicuje deklarowane wybory w znaczącym stopniu. Przejdźmy do współczynników regresji, aby zobaczyć od jakich to czynników zależy kształt tego rozkładu.

Interpretując ich wielkości w modelu dychotomicznym, traktowaliśmy je jako *odds ratio*. W modelu wielonomialnym wielkości P_i/P_6 nie mogą być interpretowane jako *odds* w ścisłym sensie, ponieważ licznik i mianownik ilorazu niekoniecznie sumują się do 1. Dla utrzymania ciągłości z modelem dychotomicznym nazwę tę można jednak zostawić z pewną modyfikacją; zwykło się je określać mianem „niewłaściwych *odds*”, czyli takich, w których mianowniku jest prawdopodobieństwo dla kategorii odniesienia, a nie odrzucenia określonego wyboru (zob. Rutherford i Choe 1993: 152).

Pozostają jednak jeszcze dwie interpretacyjne trudności. Po pierwsze, niewygodnie jest odwoływać się do pominiętej kategorii odniesienia przy określaniu siły związku pomiędzy Y i X. Zwróćmy uwagę na różnice między wartościami współczynników regresji (w tabelach 1 i 2) dla tych samych zmiennych w modelu wielonomialnym i dychotomicznym. Obydwa modele mają dostarczyć odpowiedzi na identyczne pytania, jeśli chodzi o siłę i kierunek zależności, ale wartości parametrów układają się według innego wzoru. W analizie wielonomialnej ich interpretacja jest mniej przejrzysta.

Po drugie, postępowanie się „niewłaściwymi *odds*” stwarza merytoryczny problem, ponieważ ich bezpośrednia interpretacja może prowadzić do błędnych wniosków, co uacoczni następujący przykład z tabeli 2. Jak wynikałoby z wielkości współczynnika regresji dla zmiennej „ludzie w średnim wieku”, zamieszczonego w pierwszej kolumnie, respondenci w średnim wieku rzadziej deklarowali chęć głosowania na AWS w porównaniu z najstarszymi respondentami. Różnica między obydwiema kategoriami wynosi 0,94, tj. przy wyłączeniu wpływu pozostałych zjawisk przynależność do grupy w średnim wieku zmniejsza log (P_1/P_6) o 0,94; analogicznie też zmniejsza wartość *odds*, ponieważ log (P_1/P_6) jest monotonicznie malejącą funkcją *odds* P_1/P_6 .

Stwierdzenie to jest prawdziwe dla stosunku prawdopodobieństw (logitów i *odds*), ale nie dla (P_1) — prawdopodobieństwa głosowania na AWS. Wniosek o mniejszym prawdopodobieństwie poparcia wyborczego AWS przez średnią kategorię wiekową, w porównaniu z najstarszymi wyborcami, byłby nieuzasadniony, albowiem prawdopodobieństwo wyboru AWS przez osoby w średnim wieku może przewyższać wybory deklarowane przez najstarszy elektorat. Działyoby się tak, gdyby większemu poparciu AWS przez średnią kategorię wieku (w porównaniu z najstarszą kategorią) towarzyszyły jeszcze większe sympatie wyborcze średniej grupy wieku dla Unii Pracy (względem najstarszych osób). W takim przypadku log P_1/P_6 jest wartością ujemną, a P_1/P_6 musi być mniejsze od 1. W sumie więc, log P_1/P_6 może maleć, jak ma to miejsce w przypadku rozpatrywanej zależności między średnim wiekiem a preferencjami wyborczymi dla AWS, mimo że zarówno wielkości P_1 , jak i P_6 , są dodatnie. Inaczej mówiąc ujemna wartość współczynnika regresji w modelu wielonomialnym niekoniecznie oznacza, że „przyrost” wartości zmiennej wyjaśniającej przekłada się na spadek wartości P_1 . Może być dokładnie na odwrót.

Nie mogłoby się to zdarzyć w dychotomicznej analizie logitowej, gdzie licznik i mianownik $P/(1-P)$ nie mogą się zmieniać w tym samym kierunku względem zmian wartości X_i . Wzrost P musi tu pociągać za sobą odpowiedni spadek wielkości $1-P$, w związku z czym P i $P/(1-P)$ równocześnie rosną (lub równocześnie maleją).

Przykład ten ilustruje, dlaczego w wielonomialnej analizie regresji interpretacja wpływu predyktorów X_i na log (P_i/P_j) i na P_i/P_j (na *odds*) bywa niejednoznaczna. Jest tak dlatego, że w pewnych sytuacjach zmienne X_i mogą różnicować P_i w całkiem odwrotnym kierunku niż wartości ilorazu $P_i/(1-P_i)$. Aby tych trudności uniknąć, zaleca się rezygnację z analizowania wpływu X_i na Y , wyrażonego w postaci logitów i *odds*, na rzecz bezpośredniej interpretacji: zamiast ilorazu, rozpatruje się wpływ X_i na P_i .

PRZYKŁAD BEZPOŚREDNIEJ INTERPRETACJI WPŁYWU X_i

Rozwiązanie problemu niejednoznacznej interpretacji polega na oszacowaniu wartości P_i dla wybranych wartości zmiennych niezależnych X_i . Wiedząc, jak zmieniają się P_i w zależności od X_i , można wysnuć pewne wnioski dotyczące charakteru związków zachodzących pomiędzy Y i X .

Pokażemy to na przykładzie prostego modelu, w którym są tylko 4 predyktory preferencji wyborczych, a mianowicie: podział na respondentów młodych

i w średnim wieku, poziom wykształcenia i dochody rodzin (zob. tabela 3). Ich związek z preferencjami wyborczymi można przedstawić w postaci funkcji P_i/P_6 , analogicznie jak w równaniach (3)–(7).

Tabela 3.
Zależności pomiędzy deklaracjami głosowania na AWS, SLD, PSL, UW i KPEiR w wyborach do parlamentu w 1997 r. a wybranymi cechami respondentów. Współczynniki regresji logitowej w modelu wielonomialnym

	Deklaracja głosowania na:				
	AWS	SLD	PSL	UW	KPEiR
Ludzie młodzi	0,26	0,19	-0,13	0,11	-0,22
Ludzie w średnim wieku	-0,90**	0,00	-0,47*	-0,13	-0,61**
Liczba lat nauki	0,00	0,00	0,00	-0,01	0,00
Dochody rodzin	0,00	0,00	0,00	0,00	0,00
Wyraz stały	3,79	0,93	3,20	0,23	2,51

** $p < 0,01$, * $p < 0,05$

Chi-kwadrat = 1860; Log L = -933; Liczba stopni swobody = 20

W pierwszym kroku, każde z równań (3)–(7) należy przekształcić z postaci logarytmicznej w wykładniczą, a następnie przemnożyć przez P_6 , w wyniku czego otrzymujemy:

$$P_1 = P_6 e^{a_1 + b_1 M + c_1 \dot{S} + d_1 W + f_1 I} \quad (9)$$

$$P_2 = P_6 e^{a_2 + b_2 M + c_2 \dot{S} + d_2 W + f_2 I} \quad (10)$$

$$P_3 = P_6 e^{a_3 + b_3 M + c_3 \dot{S} + d_3 W + f_3 I} \quad (11)$$

$$P_4 = P_6 e^{a_4 + b_4 M + c_4 \dot{S} + d_4 W + f_4 I} \quad (12)$$

$$P_5 = P_6 e^{a_5 + b_5 M + c_5 \dot{S} + d_5 W + f_5 I} \quad (13)$$

gdzie M i \dot{S} to „młodzi” i „średni”, W — wykształcenie, a I (duże I) — dochody rodzin na osobę.

P_6 , które występuje po prawej stronie równań (9)–(13), jest funkcją ilorazu:

$$P_6 = 1 / 1 + \sum e^{a_j + b_j M + c_j \dot{S} + d_j W + f_j I}, \quad (14)$$

Wielkość po znaku sumy w mianowniku równania (14) jest sumą odpowiednich parametrów z równań (9)–(13), gdzie j przybiera wartości od 1 do 5. Najpierw dodajemy kolejne a_1, a_2, \dots, a_5 , następnie b_1, \dots, b_5 , itd., aż do f_1, \dots, f_5 ; suma wszystkich parametrów jest wykładnikiem potęgi przy stałej e .

Po zastąpieniu P_6 w równaniach (9), (10), (11), (12) i (13), wyrażeniem występującym po prawej stronie równania (14) otrzymujemy ostatecznie 5 równań, będących podstawą estymacji P_1, \dots, P_5 dla wybranych wielkości X_i :

$$P_1 = e^{a_1 + b_1 M + c_1 \dot{S} + d_1 W + f_1 I} / 1 + \sum e^{a_j + b_j M + c_j \dot{S} + d_j W + f_j I}, \quad (15)$$

$$P_2 = e^{a_2 + b_2 M + c_2 \dot{S} + d_2 W + f_2 I} / 1 + \sum e^{a_j + b_j M + c_j \dot{S} + d_j W + f_j I}, \quad (16)$$

$$P_3 = e^{a_3 + b_3M + c_3\dot{S} + d_3W + f_3I/1} + \sum e^{a_j + b_jM + c_j\dot{S} + d_jW + f_jI} \quad (17)$$

$$P_4 = e^{a_4 + b_4M + c_4\dot{S} + d_4W + f_4I/1} + \sum e^{a_j + b_jM + c_j\dot{S} + d_jW + f_jI} \quad (18)$$

$$P_5 = e^{a_5 + b_5M + c_5\dot{S} + d_5W + f_5I/1} + \sum e^{a_j + b_jM + c_j\dot{S} + d_jW + f_jI} \quad (19)$$

Co do estymacji P_6 , to jest ona dokonywana na podstawie równania (14).

Można teraz przystąpić do wyznaczenia wartości P_1, P_2, \dots, P_6 dla wybranych X_i . Celem tej operacji jest uzyskanie prawdopodobieństwa wyboru poszczególnych partii (P_i) dla możliwie szerokiego przekroju wartości każdej ze zmiennych X ; kształt zróżnicowania przewidywanych w ten sposób preferencji wyborczych mówi coś o kierunku i sile ich zależności od X . Prawdopodobieństwa te są sumą współczynników regresji przemnożonych przez wybrane wartości odpowiadających im zmiennych. Dla przykładu pokażę, jak zmieniają się prawdopodobieństwa wyboru 6 partii ze wzrostem dochodów rodzin. W tabeli 4 podane są wielkości tych prawdopodobieństw dla dwóch sytuacji: gdy dochody na osobę były stosunkowo niskie, wynosiły 300 zł, i stosunkowo wysokie — 600 zł., dla respondentów o średniej liczbie ukończonych klas szkolnych i których wiek ustalono na poziomie odsetka najmłodszej i średniej kategorii w analizowanej próbie. Innymi słowy, są to prawdopodobieństwa dla respondentów typowych pod względem wykształcenia i wieku, ale różniących się dochodami.

Tabela 4.

Prawdopodobieństwo oddania głosu na AWS, SLD, PSL, UW, KPEiR i UP (w%) w wyborach do parlamentu w 1997 r. dla dwóch kategorii dochodów rodzin, charakteryzujących się średnimi, dla analizowanej próby: liczbą lat nauki oraz odsetkami osób młodych i w średnim wieku

Miesięczne dochody rodzin na osobę	Prawdopodobieństwo głosowania na:					
	AWS	SLD	PSL	UW	KPEiR	UP
300 zł	57,8	4,4	11,3	2,1	24,5	2,0
600 zł	60,0	4,0	4,2	0,0	43,0	3,1

Z wielkości zamieszczonych w pierwszej kolumnie tabeli 4 wynika, że prawdopodobieństwo głosowania na AWS przez osobę uzyskującą 300 zł miesięcznie wynosiło 57,8%. Zgodnie z tym, co powiedziano przed chwilą, jest to rezultat przemnożenia wartości parametrów regresji dla równania (15) przez wybrane wartości zmiennych. Występujący w liczniku równania (15) współczynnik regresji dla dochodów (-0,00036) został przemnożony przez 300, co daje -0,108; współczynniki dla trzech pozostałych zmiennych przemnożono przez ich średnie wartości (w analizowanej próbie są to: 0,25 dla zmiennej „ludzie młodzi”, 0,38 dla „ludzi w średnim wieku” i 11,1 dla poziomu wykształcenia)¹. Suma tych iloczynów plus 3,79 dla a_1 daje 3,34:

$$3,34 = 3,79 + (0,26 \cdot 0,25) + (-0,90 \cdot 0,38) + (-0,00031 \cdot 11,1) + (-0,00036 \cdot 300)$$

¹ W tym przykładzie odwołuję się do wartości współczynników regresji z pięcioma miejscami po przecinku, w tabeli 3 podane są one z zaokrągleniem do dwóch miejsc.

Następną operacją jest potęgowanie, $e^{3.34}=28,2$ — jest to licznik równania (15). Wykonując mnożenie i dodawanie odpowiednich wielkości w mianowniku, otrzymujemy 48,833. Ostatecznie, więc, 57,8% dla P_1 jest ilorazem 28,2 i 48,833.

Po wykonaniu identycznych działań liczbowych otrzymujemy prawdopodobieństwo deklarowanej chęci głosowania na AWS, w sytuacji gdy dochody rodzin równe są 600 zł. Prawdopodobieństwo to wynosi 60%. Według analogicznej formuły wyznaczono prawdopodobieństwo oddania głosów na pozostałe partie, łącznie z preferencjami wyborczymi dla Unii Pracy — podstawiając w tym ostatnim przypadku wartości zmiennych do równania (14).

Można przejść do interpretacji zależności pomiędzy analizowanymi zjawiskami, którą nazwaliśmy „bezpośrednią”. Z danych przedstawionych w tabeli 4 można wysnuć dwa podstawowe wnioski. Pierwszy z nich dotyczy niejednakowej siły prognostycznej rozpatrywanego zestawu zmiennych w zakresie przewidywania wyborów różnych partii. Ich moc prognostyczna jest zdecydowanie największa w odniesieniu do AWS, najmniejsza zaś dla Unii Wolności. Widać stąd, że poziom wykształcenia i zamożności oraz wiek stosunkowo najsilniej kształtowały zamiary związane z głosowaniem na AWS, natomiast o poparciu dla UW decydowały jakieś inne czynniki; elektorat UP, SLD i PSL jest również bardzo słabo zróżnicowany przez wiek, wykształcenie i dochody.

Drugi wniosek dotyczy kształtu zależności. Wartości zmiennej wyjaśnianej w tabeli 4 zróżnicowane są tylko pod względem dochodów i tylko na ten temat możliwe są jakieś konkluzje. Trzeba podkreślić, że postępujemy się nimi wyłącznie w celach ilustracyjnych, ponieważ prowadzenie dogłębnych interpretacji na podstawie rozkładu obejmującego jedynie dwie wartości nie ma sensu. Mając to zastrzeżenie na uwadze, można by sformułować następujący wniosek: poziom dochodów elektoratu odgrywał istotną rolę jedynie w kształtowaniu preferencji wyborczych w stosunku do PSL. Była to zależność ujemna: dwukrotnemu wzrostowi dochodów, z kwoty 300 zł miesięcznie, do 600 zł, towarzyszy prawie trzykrotny spadek prawdopodobieństwa oddania głosu na tę partię (z 11,3 do 4,2% — wyłączaając wpływ wieku i wykształcenia). Czyli PSL popierali raczej respondenci z mniej zamożnych rodzin. Warto zauważyć, że negatywny, statystycznie znaczący związek dochodów z deklaracjami sympatii dla PSL, zdaje się być potwierdzeniem wyniku regresji w modelu dychotomicznym (tabela 1), w którym wartość współczynnika dla dochodów wynosi $-0,003$ i w porównaniu z pozostałymi partiami jest to jedyna wartość znacząca.

OPROGRAMOWANIE STATYSTYCZNE

Wielonomialna analiza logitowa dostępna jest w kilku statystycznych pakietach komputerowych, z których najszerzej stosowanymi są BMDP i LIMDEP. Na zakończenie przedstawię przykład zastosowania tych analiz w pakiecie LIMDEP (w wersji 7) — podręcznik Greena (1992) zawiera szczegółową charakterystykę tego programu. Będzie to rozpatrywany powyżej model, w którym zróżnicowanie preferencji wyborczych w podziale na 6 partii wyjaśniane jest przez 4 zmienne: przynależność do kategorii osób w młodym i średnim wieku, poziom wykształcenia i dochody rodzin.

Do przeprowadzenia analiz potrzebny jest zbiór danych i słownik komend. LIMDEP nie przyjmuje danych zapisanych w formacie innych programów. Należy je przekształcić do akceptowanego formatu: jest nim „tab-delimit”, dostępny w nowszych wersjach SPSS Windows.

Po przygotowaniu zbioru można przystąpić do analiz. LIMDEP7 jest pakietem DOS. Do programu wchodzi się przez „limdep.exe”; po wejściu, na ekranie otwiera się okno „menu” — w tym momencie trzeba dokonać wyboru jednej z dwóch metod. Można wydawać kolejne polecenia bezpośrednio z klawiatury posuwając się krok po kroku („metoda interakcyjna”), lub też przygotować od razu cały słownik (w tzw. „batch system”), uruchamiając go następnie jednym ruchem. Po wykonaniu wszystkich obliczeń LIMDEP sam „wraca” do słownika. Metoda „batch” jest z kilku względów wygodniejsza, przede wszystkim dlatego, że umożliwia stałą kontrolę przebiegu wykonywanych operacji i zmniejsza ryzyko popełnienia błędu.

Słownik przygotowany w „batch system” składa się z różnych bloków komend. Pierwszy blok identyfikuje zbiór danych i zawiera definicje zmiennych — jego zadaniem jest przyłączenie słownika do danych. Następne w kolejności są polecenia rekodujące wartości zmiennych, deklaracje braków danych, tworzenie nowych zmiennych itp. Kolejny blok zawiera specyfikację modelu; po nich idą komendy nakazujące wykonanie różnych dodatkowych obliczeń. Poniżej podaję przykład słownika, którym posłużyłem się do przeprowadzenia analiz przedstawionych w tabeli 3.

```
„READ ;NVAR=5 ;NOBS=1141 ;
NAMES=WYBORY,MLODZI,SREDNI,WYKSZTAL,DOCHODY ;
FILE=CBOS97.DAT $
LIST ; WYBORY $
REJECT ; WYBORY=99 $
REJECT ; DOCHODY=9999 $
NAMELIST ; X=MLODZI,SREDNI,WYKSZTAL,DOCHODY $
CREATE ; P0=0 ; P1=0 ; P2=0 ; P3=0 ; P4=0 ; P5=0 $
NAMELIST ; PROBS=P0, P1, P2, P3, P4, P5 $
CREATE ; Y=WYBORY $
DSTAT ; RHS=Y $
LOGIT ; LHS=Y ; RHS=One,X ; MARGINAL EFFECTS $
SAVE ; FILE=DOMANSK $
ENDPROC $”
```

Znak dolara, „\$”, informuje o zakończeniu wiersza. Średniki (;) oddzielają wprowadzane informacje z wyjątkiem średnika w drugim wierszu po „names”. Komendy zamieszczone w trzech pierwszych wierszach są identyfikatorami zbioru. Otwiera je polecenie „read”, dalej podana jest liczba i nazwy zmiennych (z ograniczeniem do 8 znaków) oraz liczebność i nazwa zbioru (FILE=CBOS97.DAT). Na kolejną komendę, „LIST ; WYBORY”, LIMDEP reaguje wyświetleniem wartości zmiennej dla wszystkich kolejnych jednostek obserwacji — w naszym przypadku od 1 do 1141. Jest to opcja szczególnie przydatna przy pojawieniu się kłopotów z przyłączeniem zbioru, o czym niżej. Polecenie „REJECT” służy do eliminacji braków danych. Blok

komend od 7 do 12 wiersza (zaczynających się odpowiednio od NAMELIST i LOGIT) jest definicją analizowanego modelu. NAMELIST otwiera listę zmiennych niezależnych, CREATE i druga komenda NAMELIST specyfikują wartości zmiennej wyjaśnianej; P0 odnosi się tu do preferencji wyborczych odnośnie pominiętej w modelu Unii Pracy, a P1 do P5 identyfikują kolejno głosy na AWS, SLD, PSL, UW i KPEiR. Definicję modelu kończy komenda LOGIT z poleceniem MARGINAL EFFECTS, które mówi o tym, że LIMDEP ma dostarczyć wartości współczynników regresji i statystyk określających stopień dopasowania modelu do danych. Po czym następuje komenda SAVE nakazująca zapisanie zbioru danych i przypisanego doń słownika pod nazwą DOMANSK. Jest to wygodna opcja, przyspieszająca ponowne wejście do słownika i uruchomienie analiz. Ostatnia informacja, ENDPROC, sygnalizuje zakończenie poleceń.

Tak przygotowany słownik uruchamiany jest partiami. Najpierw trzeba go przyłączyć do danych, co w praktyce polega na zaświeceniu trzech pierwszych wierszy i wykonaniu operacji „run” wybranej z zestawu komend podanych w „menu”. Jeśli zbiór nie został przygotowany prawidłowo LIMDEP podaje informację o rodzaju popełnionego błędu. Najczęściej są to błędy wynikające z niewłaściwego formatu danych w zbiorze, np. nieakceptowalne są przecinki i kilka innych wartości kodów przyjmowanych przez SPSS. Warto w tym miejscu podkreślić, że przygotowanie dobrego zbioru zabiera w LIMDEP znacznie więcej czasu niż sporządzenie słownika i etap obliczeń.

Jeśli operacja łączenia kończy się sukcesem, można wykonać „run” dla pozostałej części słownika: w całości, blokami lub też posuwając się wiersz po wierszu. To ostatnie umożliwia sprawdzenie poprawności każdej z kolejnych komend. Podczas wykonywania poleceń zamieszczonych w ostatnim bloku, definiujących analizowany model, LIMDEP wyświetla na ekranie wszystkie wyniki testu.

LITERATURA

- Aldrich, J.H. i F.E. Nelson. 1984. *Linear Probability, Logit, and Probit Models*. London: Sage.
- Cragg, J.G. i R. Uhler. 1970. *The demand for automobiles*. „Canadian Journal of Economics” 3: 386-406.
- Demaris, A. 1992. *Logit Modeling*. London: Sage.
- Green, W.H. 1992. *LIMDEP User's Manual and Reference Guide. Version 6.0*. Bellport, N.Y.: Economic Software, Inc.
- Ellason, S.R. 1993. *Maximum Likelihood Estimation. Logic and Practice*. London: Sage.
- Heath, A. i in. 1991. *Understanding Political Change. The British Voter, 1967-1987*. Oxford: Pergamon.
- Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- Retherford, R.D. i M. Kim Choe. 1993. *Statistical Models for Causal Analysis*. New York: John Wiley & Sons.
- Schmidt, P. i R.P. Strauss. 1975. *The preconditions of occupation using multiple logit models*. „International Economic Review” 17: 204-212.
- Theil, H. 1969. *A multinomial extension of the linear logit model*. „International Economic Review” 10: 251-259.