Psychologia Społeczna
2010 tom 5 2–3 (14) 217–233
ISSN 1896-1800

Hierarchiczne modele liniowe. Co nam dają i kiedy warto je stosować

Piotr Radkiewicz[1], Marcin W. Zieliński[2]

[1] *Instytut Psychologii, Polska Akademia Nauk*
[2] *Instytut Studiów Społecznych, Uniwersytet Warszawski*

Celem artykułu jest wprowadzenie do problematyki hierarchicznych modeli liniowych – metody analitycznej zalecanej, gdy zachodzi duże prawdopodobieństwo naruszenia wymogu niezależności obserwacji. Artykuł składa się z trzech części. W pierwszej autorzy przedstawiają podstawowe metodologiczne przesłanki zastosowania metody, akcentując jej zalety w porównaniu z klasyczną analizą regresji opartą na metodzie najmniejszych kwadratów. W drugiej części omówiono najważniejsze pojęcia teoretyczne leżące u podstaw budowy modelu hierarchicznego – podział na efekty stałe i losowe, wielopoziomową strukturę danych (z uwzględnieniem interakcji międzypoziomowej) i specyficzne ujęcie składowych wariancji. Trzecia część tekstu zawiera dwa przykłady empirycznej aplikacji metody, poparte szczegółową interpretacją wyników.

Słowa kluczowe: hierarchiczne modele liniowe, niezależność obserwacji, korelacja wewnątrzklasowa, model efektów losowych, składowe wariancji

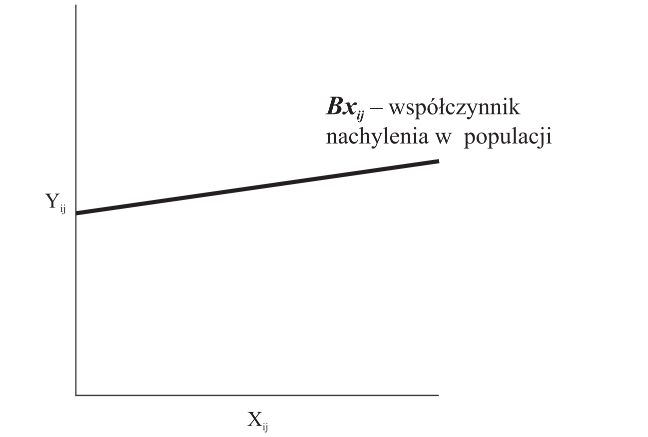
Niezależność obserwacji jest jednym z kluczowych założeń leżących u podstaw analizy regresji i innych metod analitycznych opartych na ogólnym modelu liniowym. Założenie niezależności jest spełnione, kiedy wynik pomiaru zmiennej zależnej Y, uzyskany przez dowolną osobę w zbiorze danych, nie jest zależny od wyników pomiaru zmiennej zależnej Y uzyskanych przez inne osoby z tego zbioru. W analizie regresji niezależność obserwacji stosunkowo łatwo wykazać, pokazując brak współzależności pomiędzy wynikami resztowymi (wyniki wyrażające różnicę pomiędzy obserwowanymi i przewidywanymi przez model wartościami zmiennej zależnej). Okoliczności, w których zachodzi duże prawdopodobieństwo naruszenia wymogu niezależności obserwacji, dzielone są zazwyczaj na trzy kategorie. Każda z nich odnosi się do innych źródeł potencjalnych zagrożeń.

Piotr Radkiewicz, Instytut Psychologii, Polska Akademia Nauk, ul. Chodakowska 19/31, 03–815 Warszawa, e-mail: piorad@psychpan.waw.pl
Marcin W. Zieliński, Instytut Studiów Społecznych, Uniwersytet Warszawski, ul. Stawki 5/7, 00–183 Warszawa, e-mail: marcin.zielinski@uw.edu.pl
Korespondencję w sprawie artykułu proszę kierować na adres: p.radkiewicz@uw.edu.pl.

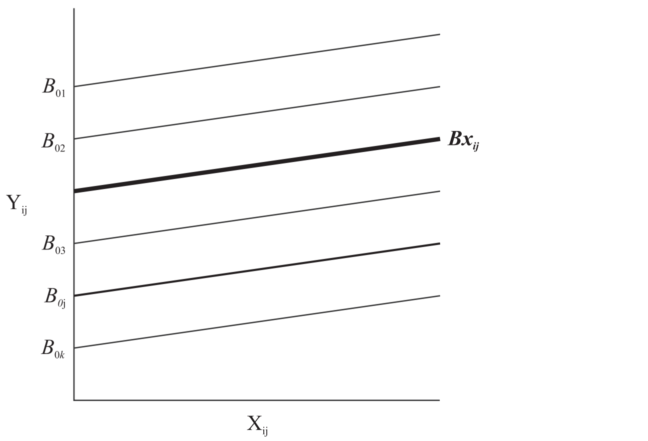
Po pierwsze, współzależności pomiędzy resztami zmiennej Y mogą być efektem systematycznych zmian w czasie. Przedmiotem zmiany są wówczas istotne właściwości uczestników badania lub samej procedury badawczej. Na przykład zdarza się, że w badaniach klinicznych pacjenci z diagnozą lekarską wskazującą na bardzo zły stan zdrowia rekrutowani są w późniejszych fazach niż pacjenci, których stan jest relatywnie dobry; w innej sytuacji niezbyt pracowici lub mało obowiązkowi studenci mogą odkładać swój udział w obligatoryjnych badaniach eksperymentalnych do samego końca semestru; może się zdarzyć i tak, że efektywność manipulacji eksperymentalnej rośnie z upływem czasu, wpływając tym samym na wariancję zmiennej zależnej.

Druga kategoria zagrożeń dotyczy tzw. zależności serii (*serial dependency*) – zjawiska charakterystycznego dla badań podłużnych, podczas których badacz dokonuje wielokrotnych pomiarów w czasie u tej samej osoby lub grupy osób. Przykładowo, mierząc każdego dnia, przez dwa miesiące, poziom stresu i nastrój w grupie osób, możemy być pewni, że pomiary nastroju, które dzieli jedna doba, będą się różniły mniej niż pomiary odległe od siebie o dwa tygodnie.

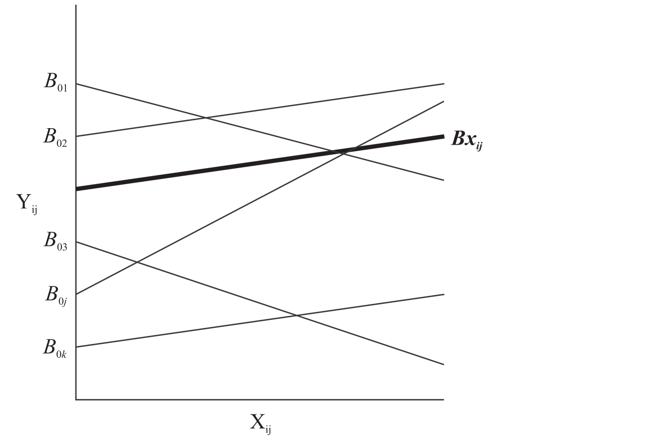
A) Regresja Y na X bez uwzględnienia grupowej struktury danych



(B) Oddzielne linie regresji dla grup. Współczynniki nachylenia w grupach są równe i nie różnią się od Bx_{ij}. Grupowe współczynniki przecięcia nie są równe (każdy z nich pokazuje poziom Y w poszczególnych grupach, gdy X = 0)



(C) Oddzielne linie regresji dla grup. Współczynniki przecięcia i nachylenia w grupach nie są równe.



Rysunek 1.
Współczynniki regresji w modelu efektów stałych i modelu efektów losowych.

Z trzecią kategorią zagrożeń mamy do czynienia, gdy materiał empiryczny zbierany jest w określonych grupach lub zbiorowościach (ang. *clustering*). Bardzo łatwo uzyskać wówczas wyniki wskazujące, że reszty regresji w obrębie grup różnią się od siebie mniej niż reszty pomiędzy grupami. Przypuśćmy, że badacz, przystępując do doboru próby respondentów, wybrał kilkadziesiąt podprób (rodziny, szkoły, uniwersytety, urzędy, lokalne wspólnoty itd.), z których następnie wylosował docelową próbę osób badanych. W tym przypadku wyniki dowolnych dwóch osób pochodzących z tej samej podgrupy będą z dużym prawdopodobieństwem bardziej do siebie zbliżone niż wyniki dwóch osób pochodzących z różnych grup. Przykładowo, w badaniach prowadzonych według takiej procedury IQ dzieci pochodzących z tej samej zbiorowości będzie prawdopodobnie mniej zróżnicowany niż IQ w próbie losowej. Podobnego efektu można oczekiwać w badaniach współmałżonków, bliźniąt itd.

Od problemu braku niezależności obserwacji nie są również wolne metody eksperymentalne. W eksperymentach polegających na śledzeniu procesów grupowych na reakcje i odpowiedzi osób badanych mogą wpływać zachowania innych uczestników badania. Zdarza się też, że w kolejnych sesjach eksperymentalnych pojawiają się niekontrolowane przez eksperymentatora różnice w procedurze badawczej, ekspozycji bodźców, zachowaniu osób przeprowadzających badanie itd.

Reasumując, komplikacje związane z niezależnością obserwacji mogą pojawić się w kilku postaciach. Te, które obejmuje kategoria pierwsza, są źródłem wariancji o nieznanej wielkości – jej kontrola w modelu analitycznym jest raczej niemożliwa. Wyeliminowaniu problemu służy przede wszystkim korekta procedury badawczej lub staranniejszy dobór próby. Problemy mieszczące się w kategorii drugiej i trzeciej doczekały się rozwiązania w postaci metody analitycznej nazywanej hierarchiczną (lub wielopoziomową) analizą danych. Ponieważ w zdecydowanej większości przypadków modele hierarchiczne są stosowane jako efektywna metoda eliminowania niepożądanych efektów grupowej struktury danych („klasteringu"), przedstawiając zasady budowy modelu hierarchicznego i kilka przykładów jego zastosowania, skoncentrujemy się na tej właśnie kategorii sytuacji badawczych. Należy jednak podkreślić, że z pewnymi modyfikacjami metoda ta jest również skutecznym środkiem kontroli wariancji pochodnej od tzw. zależności serii.

Hierarchiczny model liniowy

W klasycznej analizie regresji – opartej na metodzie najmniejszych kwadratów – istnieją trzy podejścia do zjawiska zgrupowania obserwacji i jego pochodnych.

W pierwszym, nazywanym dezagregacją, problem jest całkowicie ignorowany, a analiza przebiega tak, jakby dane nie miały struktury grupowej. Drugie podejście polega na agregacji danych na poziomie grupowym. Otrzymujemy średnie grupowe wszystkich predyktorów i zmiennej zależnej, a grupy traktowane są jako jednostki analizy. W rezultacie, równanie regresji opisuje związek pomiędzy średnimi wartościami predyktorów w grupach i średnimi grupowymi zmiennej zależnej. W literaturze przedmiotu podkreśla się, że inferowanie z poziomu równania dla danych zgrupowanych o właściwościach pojedynczych obserwacji może prowadzić do całkowicie błędnych wniosków (np. Cohen, Cohen, West i Aiken, 2003).

Zastosowanie dezagregacji albo agregacji skutkuje estymacją równań z zupełnie innymi współczynnikami regresji. O ile w pierwszym wypadku można je nazwać ogólnymi, ponieważ opisują zależności w populacji składającej się z pojedynczych obserwacji, o tyle w drugim, kiedy obserwacjami są średnie wartości grupowe, współczynniki regresji *de facto* odzwierciedlają międzygrupowe różnice dla efektu predyktora na zmienną zależną. Choć w sensie matematycznym obie metody niczym się nie różnią (co ilustruje Rysunek 1a), bywa, że opisują całkowicie inne wymiary społecznej rzeczywistości. Na przykład, na poziomie danych zagregowanych wysoki poziom indywidualizmu, obserwowany w poszczególnych krajach Europy, wyraźnie sprzyja zaangażowaniu obywateli w sprawy publiczne i inicjatywy społeczne, natomiast na poziomie danych jednostkowych dodatni związek pomiędzy indywidualizmem i aktywnością społeczną praktycznie nie istnieje, a nawet przyjmuje postać zależności negatywnej.

Model analizy regresji oparty na metodzie najmniejszych kwadratów pozwala też na jeszcze inne, trzecie podejście do problemu danych zgrupowanych. Polega ono na regresji zmiennej zależnej na predyktory reprezentujące indywidualny poziom danych, z jednoczesnym włączeniem do modelu, w celu identyfikacji grupowej przynależności obserwacji, zbioru $g - 1$ (g = liczba grup) tzw. zmiennych zero-jedynkowych (*dummy codes*). Po dokonaniu takiego zabiegu współczynnik nachylenia dla predyktora X odzwierciedla efekt łącznej regresji wewnątrzklasowej (*pooled within-class regression*), który interpretuje się jako ważoną przeciętną wartość współczynnika regresji ze wszystkich grup obserwacji.

Istotę kodowania zmiennych zero-jedynkowych ilustruje Rysunek 1b. Podejście to umożliwia śledzenie efektu predyktora na zmienną zależną po usunięciu różnic wynikających ze średnich grupowych. Do tego momentu przyjmujemy, że siła związku między X i Y wyrażona współczynnikiem B (nachylenia) ma stałą wielkość, a linie regresji wyznaczone dla poszczególnych grup różnią się jedynie wartością stałej (współczynnik przecięcia), co wynika z różnic pomiędzy średnimi grupowymi zmiennej Y. Metoda kodowania zmiennych zero-jedynkowych pozwala jednak na jeszcze więcej: poprzez dołączenie do równania regresji dodatkowych zmiennych instrumentalnych umożliwia modelowanie zależności interakcyjnych pomiędzy zmienną grupującą i predyktorem. Przedstawia to rysunek 1c, na którym widzimy, że linie regresji Y na X, wyznaczone dla poszczególnych grup, różnią się nie tylko wielkością stałej, lecz także wielkością (a nawet kierunkiem) współczynnika B.

Metodę eliminacji problemu „klasteringu”polegającą na kodowaniu zero-jedynkowym, czyli wychwytującym tę część wariancji zmiennej zależnej, za którą odpowiadają różnice międzygrupowe, nazywa się modelem efektów stałych. Jest to zgodne z nomenklaturą, według której klasyczna analiza regresji (oparta na metodzie najmniejszych kwadratów) określana jest jako model regresji dla efektów stałych (*fixed effects coefficient regression model)*.

Jak argumentowaliśmy powyżej, kiedy dane mają strukturę zgrupowaną („klastering”), klasyczna analiza regresji może doprowadzić badacza do całkowicie błędnych wniosków. Wydaje się, że obecnie najpopularniejszą i najbardziej efektywną alternatywą jest model regresji dla efektów losowych (*random effects coefficient regression model)*. Różni się on od klasycznego pewnymi założeniami dotyczącymi współczynników regresji i struktury skorelowania pojedynczych obserwacji. Ponadto podejście zaproponowane w modelu efektów losowych umożliwia pomiar zmiennych na kilku poziomach analizy – jednostkowym i grupowym. Na przykład, w badaniach międzykulturowych jednostkowym wskaźnikiem sytuacji materialnej mogą być dochody respondenta, a na poziomie grupowym wielkość dochodu narodowego *per capita*. Modele regresji dla efektów losowych pozwalające na wielopoziomową analizę danych nazywa się najczęściej hierarchicznymi modelami liniowymi (*hierarchical linear modeling*) (np. Bryk i Raudenbush, 1992; Goldstein, 1995; Hox, 2002; Kreft i de Leeuw, 1998).

Korelacja wewnątrzklasowa

W tradycyjnej analizie regresji zgrupowanie danych skutkuje zazwyczaj zniekształceniem błędów standardowych współczynników regresji (najczęściej są one zaniżone). W rezultacie przedziały ufności wyznaczane dla współczynników są zbyt wąskie, co powoduje przeszacowanie wartości testów istotności statystycznej. Im większa homogeniczność wyników w obrębie grup (efekt „klasteringu”), tym więcej obaw, że pojawi się zjawisko inflacji poziomu alfa. Mamy z nim do czynienia wtedy, gdy rzeczywisty poziom błędu pierwszego rodzaju (prawdopodobieństwo odrzucenia H_0, gdy jest prawdziwa) przekracza nominalny, konwencjonalnie przyjęty poziom krytyczny (czyli najczęściej 0,01 lub 0,05).

Poziom zgrupowania danych mierzy się najczęściej współczynnikiem korelacji wewnątrzklasowej (ICC od ang. *intra-class correlation*; Bryk i Raudenbush,1992; Hox, 2002). Pokazuje on proporcję całkowitej wariancji zmiennej zależnej wyjaśnianą przez przynależność grupową obserwacji. Na ICC można również patrzeć jak na miarę stopnia podobieństwa członków tej samej kategorii grupowej (na ile są oni bardziej podobni do siebie niż do członków innych kategorii grupowych). Współczynnik ICC obliczany jest ze wzoru:

$$(1) \qquad ICC = \left(\frac{\tau_{00}}{\tau_{00} + \sigma^2} \right)$$

gdzie τ_{00} oznacza wariancję międzygrupową, a σ^2 wariancję wewnątrz grup.

W celu oszacowania korelacji wewnątrzklasowej wykorzystuje się wartości zmiennej zależnej pobrane z poszczególnych grup. Mogą to być na przykład wyniki testu kompetencji szkolnych zebrane w kilkudziesięciu szkołach. Do obliczenia ICC wystarczy wówczas prosty model jednoczynnikowej analizy wariancji dla jednokrotnych pomiarów (ANOVA), w którym czynnikiem jest zmienna grupująca (szkoła), a zmienną zależną wynik testu kompetencji.

Zakres zmienności ICC waha się od 0 (całkowita niezależność) do 1 (pełna współzależność obserwacji). Jeśli grupy nie różnią się od siebie, ICC = 0 i jest to zgodne z założeniami ogólnego modelu liniowego. Nawet przy niewielkiej wartości ICC, takiej jak 0,01 czy 0,05, rzeczywisty poziom α, na którym jest odrzucana hipoteza zerowa, wzrasta dramatycznie. Przykładowo, w jednoczynnikowej analizie wariancji z równolicznymi grupami $n = 25$, przy ICC = 0,01, rzeczywisty poziom α dla testu istotności statystycznej efektu manipulacji wynosi 0,11, podczas gdy nominalnie, zgodnie z tablicami statystycznymi, krytycznej wartości testu odpowiada α = 0,05. Przy tych samych liczebnościach grupowych i ICC = 0,05 rzeczywisty poziom α wzrasta do 0,19. Inflacja wartości α rośnie wraz ze wzrostem współczynnika ICC i liczebności próby. Podobny mechanizm obserwujemy w przypadku analizy regresji, w której badacz wykorzystuje dane zgrupowane.

Wielopoziomowa struktura modelu losowych współczynników regresji (efektów losowych)[1]

Model efektów losowych jest znacznie bardziej złożony niż model regresji oparty na metodzie najmniejszych kwadratów, ponieważ jako immanentne składniki analizy uwzględnia grupowaną strukturę danych oraz indywidualny i grupowy poziom relacji między zmiennymi. Istnieją w nim trzy typy równań: 1) poziom pierwszy (mikro) – jedno równanie dla każdej grupy w zbiorze danych; 2) poziom drugi (makro)[2] – opisujący grupową, niekiedy wielopoziomową strukturę danych; i 3) model mieszany (*the mixed model equation*) – równanie dla połączonych efektów pierwszego i drugiego poziomu. W omówieniu przedstawionym poniżej, dla uproszczenia i przejrzystości, model zawiera tylko jeden predyktor na każdym poziomie. Należy jednak pamiętać, że zarówno na poziomie mikro, jak i makro możemy umieścić dowolną liczbę predyktorów.

Równanie regresji na poziomie mikro dla dowolnej grupy wygląda następująco (objaśnienia dotyczące notacji zastosowanej w kolejnych równaniach znajdują się w aneksie umieszczonym na końcu tekstu):

$$(2) \qquad \mathrm{y}_{\mathrm{ij}} = B_{1j}\mathrm{x}_{ij} + B_{0j} + r_{ij}$$

Ponieważ zakładamy, że obserwowane grupy stanowią losową próbę z populacji wszystkich grup, stałe i współczynniki B z poziomu 1 stają się zmiennymi losowymi w modelu efektów losowych. Inaczej mówiąc, każdy model hierarchiczny generuje całą serię analiz regresji, jedną dla każdej grupy, z ich własnymi współczynnikami. W ten sposób pojedyncza analiza regresji dla efektów losowych opiera się na rozkładach grupowych współczynników B_{1j} i B_{0j}, zgodnie z założeniem, że B_{1j} i B_{0j} przyjmują postać zmiennych losowych.

W modelu analizy regresji dla efektów losowych wnioskujemy o współczynnikach z populacji. Równania z poziomu makro obrazują relacje grupowych estymatorów B_{0j} i B_{1j} z ich odpowiednikami dla całej populacji estymowanymi w modelu efektów stałych. Zgodnie z przyjętą notacją zakładamy, że wartości stałej B_{0j} z poziomu 1 dla obserwowanych grup tworzą rozkład losowy wokół stałej w populacji Υ_{00}, a wartości współczynników B_{1j} z poziomu 1 dla tych samych grup mają losowy rozkład wokół populacyjnego współczynnika nachylenia Υ_{10}. Ilustrują to równania dla poziomu makro:

(3) *poziom 2 dla współczynnika przecięcia*
$$B_{0j} = \Upsilon_{00} + u_{0j}$$

(4) *poziom 2 dla współczynnika nachylenia*
$$B_{1j} = \Upsilon_{10} + u_{1j}$$

Z równania (3) wynika, że grupowa wartość B_{0j} jest addytywnym efektem, na który składają się wartość stałej w populacji (γ_{00}) i losowy element u_{0j} przedstawiający odchylenie B_{0j} od Υ_{00} ($u_{0j} = B_{0j} - \Upsilon_{00}$). Analogicznie, w równaniu (4) postulujemy, że grupowa wartość współczynnika B_{1j} jest addytywnym efektem, na który składają się wartość współczynnika nachylenia w populacji (Υ_{10}) i losowy element u_{1j} wyrażający odchylenie B_{1j} od Υ_{10} ($u_{1j} = B_{1j} - \Upsilon_{10}$). W ten sposób, poprzez dodanie poziomu makro, bierzemy pod uwagę grupową strukturę zmiennych. Jeśli wariancja współczynników grupowych wynosi 0, obydwa modele – dla efektów losowych i dla efektów stałych – są równoważne.

Potraktowane łącznie, dwa poziomy analizy opisane powyżej tworzą tzw. model mieszany. Po podstawieniu współczynników z wzorów (3) i (4) do wzoru z poziomu 1 (mikro) otrzymujemy następujące równanie:

$$
\begin{aligned}
& y_{ij} = (\Upsilon_{00} + u_{0j}) + (\Upsilon_{10} + u_{1j})\, x_{ij} + r_{ij} \\
(5)\quad & y_{ij} = \Upsilon_{00} + u_{0j} + \Upsilon_{10}x_{ij} + u_{1j}x_{ij} + r_{ij} \\
& y_{ij} = \Upsilon_{00} + \Upsilon_{10}x_{ij} + (u_{0j} + u_{1j}x_{ij} + r_{ij})
\end{aligned}
$$

Równanie dla modelu mieszanego (5) przedstawia regresję zmiennej zależnej Y na predyktor z poziomu 1 w postaci właściwej współczynnikom w populacji. Wariancja reszt regresji ma charakter addytywny, łącząc składniki błędu pochodzące z obu poziomów analizy – mikro (r_{ij}) i makro (u_{0j} i $u_{1j}x_{ij}$). Wartości reszt r_{ij} interpretujemy tak samo, jak w klasycznej analizie regresji – jako różnicę pomiędzy obserwowaną i przewidywaną wartością zmiennej zależnej Y. Wartości u_{0j} i u_{1j} pokazują różnice pomiędzy grupowymi i populacyjnymi wielkościami współczynników, odpowiednio, przecięcia i nachylenia. Wielkość błędu w modelu mieszanym jest większa niż odpowiadający jej błąd w zdezagregowanej analizie regresji, która ignoruje grupową przynależność obserwacji.

Współczynniki Υ_{00} i $\Upsilon_{10}x_{ij}$, otrzymane w wyniku analizy regresji dla efektów losowych pojedynczego predyktora x, mogą mieć wielkości bardzo zbliżone do współczynników B_0 i B_1 uzyskanych w klasycznej analizie regresji dla efektów stałych. Należy się jednak spodziewać, że błąd standardowy towarzyszący Υ_{00} i $\Upsilon_{10}x_{ij}$ będzie większy, niż błąd standardowy towarzyszący ich odpowiednikom w zdezagregowanej analizie regresji. Różnica ta będzie rosła proporcjonalnie do wzrostu wariancji grupowych B_{0j} i B_{1j}.

Na zakończenie tego paragrafu warto zauważyć, że możliwa jest sytuacja, gdy wariancja efektów losowych wynosi 0. Zdarza się na przykład, że współczynniki przecięcia pokazują wyraźne różnice międzygrupowe, natomiast nachylenie współczynników B zachowuje międzygrupową stałość (por. Rysunek 1b). Wariancję stałych można wówczas traktować jako efekt losowy, a wariancję współczynników B jako efekt stały. Model taki jest tożsamy z klasyczną analizą regresji dla efektów stałych, w której różnice pomiędzy średnimi grupowymi reprezentują wektory zmiennych instrumentalnych.

Predyktory na poziomie makro oraz interakcja predyktorów poziomów 1 i 2

Do tej pory można było odnieść wrażenie, że „klastering” jest jedynie uciążliwym problemem metodologicznym, komplikującym badaczowi ocenę rzeczywistego związku predyktora z poziomu 1 i zmiennej zależnej. Nic bardziej mylnego. W wielu wypadkach zgrupowana struktura danych jest sama w sobie źródłem interesujących pytań badawczych. A bywa i tak, że stanowi istotę problemu badawczego. Na przykład badacz, którego interesuje ocena wpływu indywidualnych zdolności ucznia i efektywności kształcenia na wyniki w teście osiągnięć szkolnych, może przeprowadzić badanie, w którym dokonuje pomiaru IQ (predyktor na poziomie 1) w kilkudziesięciu szkołach (zmienna grupująca), a następnie do analiz dołączą wyniki zewnętrznego rankingu jakości kształcenia szkolnego (predyktor na poziomie 2), będące syntetycznym wskaźnikiem opartym na zobiektywizowanych kryteriach oceny (średni wynik w ogólnokrajowym teście kompetencji, odsetek kandydatów przyjętych na studia wyższe, liczba uczniów biorących udział w olimpiadach przedmiotowych itd.). W takiej sytuacji wartości predyktora z poziomu 2 (wyniki w rankingu) nie są przypisywane pojedynczym obserwacjom, lecz wyłącznie jednostkom grupującym (tu: szkoły).

Jeśli problem badawczy uwzględnia działanie predyktora z poziomu 2 (makro), konieczne jest poszerzenie hierarchicznego modelu regresji o dodatkowe elementy. Nie zmienia to oczywiście równania na poziomie 1, natomiast na poziomie 2 wygląda ono następująco:

(6) *poziom 2 dla współczynnika przecięcia*
$$B_{0j} = \Upsilon_{01}\, W_j + \Upsilon_{00} + u_{0j}$$

(7) *poziom 2 dla współczynnika nachylenia*
$$B_{1j} = \Upsilon_{11}\, W_j + \Upsilon_{10} + u_{1j}.$$

Równania (6) i (7) przedstawiają efekt dołączenia predyktora W*j* na poziomie makro, kiedy badacz przypuszcza, że grupowe współczynniki przecięcia i/lub nachylenia zależą od wariancji predykatora z poziomu 2. Jeśli równanie (6) odniesiemy do przykładu badań w szkołach, będzie ono opisywało hipotezę mówiącą, że średni przyrost wyników w teście osiągnięć zależy od jakości kształcenia (efekt $\Upsilon_{01}W_j$). Badacz oczekuje, że wysokie wyniki w rankingu jakości kształcenia pozwalają przewidywać wyższy średni przyrost wyników testu osiągnięć w szkołach. Z kolei w równaniu (7) pojawia się domniemanie, że związek pomiędzy poziomem IQ a wynikiem w teście osiągnięć zależy od jakości kształcenia (efekt $\Upsilon_{11}W_j$). Badacz spodziewa się synergicznego efektu interakcji, w którym siła tego związku rośnie wraz ze wzrostem pozycji szkoły w rankingu.

Model poszerzył się o dwa nowe efekty stałe – Υ_{01} i Υ_{11}. Pierwszy z nich reprezentuje grupowy współczynnik przecięcia B_{0j}, a drugi grupowy współczynnik nachylenia B_{1j} dla predyktora poziomu makro W_j. Pojawienie się predyktora W_j na poziomie makro zmienia postać zapisu modelu mieszanego. Po podstawieniu współczynników z wzorów (6) i (7) do wzoru z poziomu 1 (mikro) otrzymujemy równanie:

$$\begin{aligned}
y_{ij} &= (\Upsilon_{01} W_j + \Upsilon_{00} + u_{0j}) + \\
&\quad + (\Upsilon_{11} W_j + \Upsilon_{10} + u_{1j})\, x_{ij} + r_{ij} \\
y_{ij} &= \Upsilon_{01} W_j + \Upsilon_{00} + u_{0j} + \Upsilon_{11} W_j x_{ij} + \\
&\quad + \Upsilon_{10} x_{ij} + u_{1j}\, x_{ij} + r_{ij} \\
y_{ij} &= \Upsilon_{01} W_j + \Upsilon_{10} x_{ij} + \Upsilon_{11} W_j x_{ij} + \\
&\quad + \Upsilon_{00} + (u_{0j} + u_{1j}\, x_{ij} + r_{ij})
\end{aligned} \tag{8}$$

Trzy pierwsze elementy w równaniu (8) przedstawiają, kolejno, efekty jakości kształcenia (predyktor poziomu 2), poziomu zdolności (predyktor poziomu 1) i „międzypoziomowej" interakcji obu predyktorów (*cross-level interaction*). Interakcja pomiędzy poziomami wskazuje, że zmienna W_j z poziomu makro jest predyktorem wielkości współczynnika nachylenia na poziomie mikro, czyli moderuje związek IQ i wyniku w teście osiągnięć szkolnych. W modelu znajdują się teraz cztery efekty stałe – Υ_{01}, Υ_{10}, Υ_{11} i Υ_{00}, odpowiadające, kolejno, współczynnikom w klasycznym równaniu regresji z dwoma predyktorami i efektem interakcji – $Y = B_1W + B_2X + B_3WX + B_0$.

Składowe wariancji

Model analizy regresji z efektami losowymi zawiera jeszcze jeden charakterystyczny element, którego nie znajdziemy w klasycznej analizie regresji. Są to tzw. składowe wariancji. W modelach hierarchicznych wyodrębnia się trzy niezależne źródła losowej wariancji błędu: 1) wartości reszt r_{ij} zmiennej zależnej Y z poziomu 1; 2) wartość u_{0j} pokazującą różnice pomiędzy grupowymi i populacyjnymi wielkościami współczynników przecięcia; oraz 3) wartość u_{1j} określającą ilościowe różnice pomiędzy grupowymi i populacyjnymi wielkościami współczynników nachylenia. Stanowią one trzy składowe wariancji, indeksowane jako σ^2, τ_{00} i τ_{11}.

Składowe τ_{00} i τ_{11} opisują wpływ zgrupowanej struktury danych na związek predyktora i zmiennej zależnej. Badacz nie ignoruje istnienia różnic międzygrupowych (co zakłada dezagregacja) ani nie redukuje w sposób sztuczny wariancji wewnątrzgrupowej (efekt agregacji). Taka forma konceptualizacji wariancji błędu ma zasadniczą przewagę nad modelem regresji dla efektów stałych, w którym „klastering" znajduje odzwierciedlenie w postaci zmiennych instrumentalnych. Zwróćmy uwagę, że podczas gdy model regresji z efektami losowymi wymaga jedynie dwóch składowych wariancji (τ_{00} i τ_{11}) opisujących grupowe różnice współczynników przecięcia i nachylenia, w modelu ze zmiennymi instrumentalnymi potrzebujemy zbioru $g - 1$ zmiennych opisujących wariancję współczynników przecięcia i drugiego zbioru $g - 1$ zmiennych kodujących interakcje poziomu grupowego z predykatorem poziomu 1, czyli wariancję współczynników nachylenia. Zatem dwie składowe wariancji z powodzeniem zastępują $2(g - 1)$ zmiennych instrumentalnych.

Modele hierarchiczne mają jeszcze jedną składową wariancji. Jest nią kowariancja losowych współczynników przecięcia i nachylenia w grupach (poziom 1). Współzależność obu efektów może być bardzo interesująca z teoretycznego punktu widzenia. W przykładzie szkolnym, jeśli związek ten jest dodatni, oznaczać to będzie, że w szkołach osiągających najlepsze wyniki nauczania istnieje najsilniejszy związek pomiędzy IQ i testem osiągnięć. W przypadku związku ujemnego, w najlepszych szkołach związek ten byłby najmniejszy. Związek pozytywny mógłby więc być argumentem na rzecz hipotezy, że dobra szkoła pozwala na pełniejszą ekspresję naturalnych zdolności ucznia; związek negatywny dowodziłby z kolei, że naturalne zdolności mają największe znaczenie wtedy, gdy pozwalają uczniowi kompensować niedostatki kształcenia szkolnego.

Na koniec wróćmy na chwilę do równania (8) i zawartego w nim całościowego składnika błędu. Znajdziemy w nim te same elementy, które pojawiają się w równaniu (5). Z jednym wszakże wyjątkiem – w równaniu (5) nie uwzględnialiśmy efektów związanych z wprowadzeniem predyktora na poziomie makro. Jego dodanie zmienia interpretację obu składników błędu – u_{0j} i u_{1j}. W tej chwili efekt u_{0j} wyraża tę część różnicy pomiędzy współczynnikiem przecięcia w grupie j i w populacji, której nie można wyjaśnić działaniem predyktora z poziomu 2. Jeśli predyktor W wiąże część zmienności współczynnika przecięcia, wartość u_{0j} będzie mniejsza niż w modelu bez predyktora W. W przykładzie szkolnym u_{0j} reprezentuje tę część średniego grupowego wyniku w teście osiągnięć, której nie można wyjaśnić efektem jakości kształcenia (pozycja szkoły w rankingu). To samo dotyczy składnika u_{1j} – jeśli predyktor z poziomu 2 wyjaśnia pewną część zmienności współczynników nachylenia, u_{1j} będzie mniejszy niż w modelu bez tego predyktora.

Podsumowując ostatni paragraf, możemy skonkludować: modelowanie hierarchiczne pozwala wyjaśniać wariancję losowych efektów współczynników przecięcia i nachylenia działaniem predyktorów z poziomu 2 (makro). Rzecz więc nie tylko w tym, aby poddać statystycznej kontroli wariancję będącą pochodną różnic pomiędzy grupowymi współczynnikami regresji. Chodzi również o to, aby wskazać jej przyczyny.

Estymacja parametrów

Sposób estymowania parametrów to jedna z kluczowych różnic pomiędzy klasycznym modelem efektów stałych i modelem efektów losowych. W pierwszym estymatory współczynników w populacji otrzymujemy dzięki doskonale znanej metodzie zwykłych najmniejszych kwadratów, natomiast w modelu efektów losowych estymacji parametrów dokonuje się za pomocą tzw. metody największej wiarygodności (*maximum likelihood estimation*, ML) lub blisko z nią powiązanej metody ograniczonej największej wiarygodności (REML). Obie one opierają się na iteracyjnej procedurze szacowania efektów stałych i składowych wariancji. Procedura ta rozpoczyna się od zdefiniowana początkowego zbioru parametrów próby (współczynników i ich błędów standardowych), które służą następnie za podstawę estymacji kolejnego zbioru parametrów zastępujących zbiór wartości początkowych. Jest to pierwszy krok iteracji. Po nim następują kolejne, aż do momentu, kiedy procedura osiągnie tzw. konwergencję, tj. wartości parametrów otrzymanych w którymś kolejnym kroku różnią się od otrzymanych w kroku poprzednim tak niewiele, że wielkość tej różnicy nie przekracza przyjętej wartości kryterialnej (tzw. kryterium konwergencji).

Przykłady zastosowania hierarchicznej analizy danych: badanie międzykulturowe i badanie eksperymentalne

Czy modele hierarchiczne można wykorzystać w badaniach psychologicznych? Oczywiście tak. Choć z pewnością nie jest to metoda, po którą badacze sięgają często, jej popularność z roku na rok rośnie. Na jej relatywną słabą znajomość składa się kilka przyczyn. Po pierwsze, warto pamiętać, że modele hierarchicznej analizy danych pojawiły się w naukach społecznych stosunkowo niedawno – pierwsze prace przedstawiające matematyczne podwaliny modelu i jego zastosowanie pochodzą z pierwszej połowy lat 90. ubiegłego wieku (Bryk i Raudenbush, 1992; Hox, 1995), a zatem od ich wydania upłynęło raptem kilkanaście lat. Po drugie, programy komputerowe opracowane dla modelowania hierarchicznego dość długo były mało dostępne, a większość najpopularniejszych pakietów statystycznych do dziś nie zawiera specjalnych modułów przeznaczonych do tego typu analiz. Po trzecie wreszcie, trzeba otwarcie powiedzieć, że pytania i problemy badawcze, dla których modele hierarchiczne są najbardziej użytecznym narzędziem, ani w warstwie teoretyczno-poznawczej, ani metodologicznej, nie należą do tzw. głównego nurtu szeroko rozumianej psychologii społecznej.

W dalszej części artykułu przedstawimy dwa przykłady badań psychologicznych, w których zastosowanie modeli hierarchicznych jest szczególnie zalecane. Pierwszy z nich dotyczył będzie obszaru tzw. psychologii międzykulturowej, ponieważ wydaje się, że właśnie w tym nurcie badań psychologicznych modele hierarchiczne cieszą się największą popularnością. Klasyczny problem badawczy polega tu na wykazaniu wpływu zgeneralizowanych, charakterystycznych dla całych społeczeństw orientacji kulturowych (zmienna makro z poziomu 2) na indywidualne charakterystyki osób badanych (zmienna mikro z poziomu 1), opisujące właściwości osobowości, percepcję społeczną, postawy itd. (np. Berry, Poortinga, Segall i Dasen, 2002). W badaniach z tego obszaru cel może być też inny: badacze mogą skoncentrować się na pytaniu o relatywne znaczenie czynników indywidualnych i makrospołecznych w wyjaśnieniu określonych postaw i zachowań (np. Coenders, Lubbers i Scheepers, 2007). W tym przypadku może się okazać, że szczególnie ciekawy jest nie tyle efekt interakcji między poziomami mikro i makro, ile na przykład fakt, że zachowania polegające na dyskryminacji cudzoziemców w większym stopniu zależą od poziomu dochodów całego społeczeństwa (zmienna z poziomu 2) niż od indywidualnych dochodów osoby badanej (zmienna z poziomu 1).

W przykładzie drugim odwołamy się do badań eksperymentalnych. Cohen, Cohen, West i Aiken (2003) opisują pewną kategorię eksperymentów, które są poświęcone efektom interakcji cech osobowości jednostki (wyrażających różnice indywidualne) i manipulacji eksperymentalnej. Celem takich badań jest odpowiedź na pytanie, do jakiego stopnia reakcja osoby badanej na manipulację eksperymentalną jest funkcją stałych indywidualnych charakterystyk osobowości, którymi ludzie różnią się między sobą. W analizach hierarchicznych manipulację eksperymentalną i pomiar osobowości stosunkowo łatwo umieścić w modelu, w którym udział w jednym z warunków eksperymentalnych (grupa kontrolna *vs.* eksperymentalna) traktujemy jako predyktor z poziomu 2, a indywidualne charakterystyki osobowości, mogące wyznaczać reakcje badanego na manipulację eksperymentalną, jako predykator z poziomu 1. Oczywiście kiedy badacz ma pewność, że wyniki nie są zależne od grupowej struktury danych, może w analizach zastosować klasyczny model regresji. Jednak w wielu badaniach prowadzonych w schemacie eksperymentalnym możemy napotkać problem „klasteringu", którego obecność oznacza, że struktura próby lub zachodzące w niej procesy grupowe (na przykład wzrost wewnętrznej spójności grupy) mogą wpływać na wariancję zmiennej zależnej. Wtedy właśnie warto rozważyć zastosowanie modelu hierarchicznej analizy danych. W dalszej części tekstu, jako przykład drugi, wykorzystamy opis jednego z takich badań wykonanego w schemacie eksperymentalnym z pomiarem zmiennej osobowościowej (*experimental personality design*; West, Aiken i Krull, 1996).

W przykładzie pierwszym analizy hierarchiczne zostały wykonane za pomocą pakietu Nlme (Pinheiro, Bates, DebRoy i Sarkar, 2009), będącego jedną ze składowych bezpłatnego programu statystycznego R (R Development Core Team, 2009). Wyniki dla modelu hierarchicznego przedstawionego w przykładzie drugim obliczono w pakiecie statystycznym SAS z użyciem modułu SAS PROC MIXED[3].

Przykład 1 – badanie międzykulturowe

Pierwszy przykład zastosowania modelu analizy hierarchicznej został zaczerpnięty z archiwów *International Social Survey Programme* (ISSP), programu badawczego realizowanego od 1985 roku w kilkudziesięciu krajach Europy, Ameryki i Azji. W ramach ISSP poszczególne kraje członkowskie zobowiązane są do corocznej realizacji sondażu dotyczącego zaakceptowanej przez wszystkich uczestników tematyki badawczej, z wykorzystaniem ujednoliconego narzędzia badawczego (kwestionariusz) i wystandaryzowanych metod realizacji badania. Dane wykorzystane w analizach pochodzą z modułu poświęconego tożsamości narodowej (*National Identity I*), w którego realizacji (1995 rok) uczestniczyły 23 kraje członkowskie.

Wykorzystując modelowanie hierarchiczne, badacze chcieli odpowiedzieć na kilka pytań. Po pierwsze, przedmiotem ich zainteresowania na poziomie mikro była tożsamość narodowa i jej związki z ksenofobią. Sądzili, że osoby wykazujące inkluzywny (włączający, otwarty) schemat narodowej tożsamości, oparty na małej liczbie możliwych do spełnienia kryteriów, będą przejawiały znacznie niższy poziom ksenofobii niż osoby posługujące się schematem ekskluzywnym (wykluczającym, zamkniętym). Ponadto badacze chcieli sprawdzić, czy obserwowane pomiędzy krajami różnice w nasileniu ksenofobii można wyjaśnić poziomem rozwoju ekonomicznego społeczeństwa (zmienna makro) oraz czy rozwój ekonomiczny – mierzony wskaźnikiem PKB *per capita* – moderuje na poziomie mikro siłę związku między inkluzywnością tożsamości narodowej i ksenofobią. Przyjęto hipotezę badawczą o interakcji, zakładając, że negatywny związek inkluzywnej tożsamości narodowej i ksenofobii będzie najsilniejszy w krajach cieszących się najwyższym poziomem PKB *per capita* a najsłabszy w krajach najbiedniejszych.

Do pomiaru ksenofobii (zmienna zależna) badacze użyli narzędzia składającego się z czterech pozycji określających stosunek respondenta do imigrantów z innych krajów (przykłady: „Przybysze z innych krajów przyczyniają się do zwiększenia przestępczości", „Przybysze z innych krajów przyczyniają się do rozwoju polskiej gospodarki"). Respondenci odpowiadali na skali od 1 (zdecydowanie się zgadzam) do 5 (zdecydowanie się nie zgadzam). Po zrekodowaniu dwóch pozycji i uśrednieniu wszystkich odpowiedzi wysoki wynik ogólny wskazywał na silną ksenofobię. Rzetelność wewnętrzna narzędzia mierzona współczynnikiem alfa Cronbacha wynosiła dla całego zbioru danych 0,86 (od 0,78 do 0,90 w poszczególnych próbach krajowych).

Predyktorem na poziomie indywidualnym była tożsamość narodowa na wymiarze ekskluzywność *vs.* inkluzywność. Do pomiaru użyto siedmiu cech kryterialnych, którymi powinna charakteryzować się osoba będąca „prawdziwym Polakiem" (Niemcem, Czechem, Hiszpanem itd.) (np. być urodzonym w Polsce, mieć polskie obywatelstwo, mieszkać w Polsce przez większą część życia). Respondenci odpowiadali na skali od 1 (bardzo ważne) do 4 (w ogóle nieważne). Po odwróceniu i uśrednieniu wszystkich odpowiedzi, wysoki ogólny wynik wskazywał na inkluzywność (otwartość) tożsamości narodowej. Rzetelność wewnętrzna narzędzia mierzona współczynnikiem alfa Cronbacha wynosiła dla całego zbioru danych 0,82 (od 0,75 do 0,86 w poszczególnych próbach krajowych).

Na poziomie grupującym pojawił się jeden predyktor – produkt krajowy brutto (PKB) *per capita*. Informacje o PKB *per capita* w dolarach amerykańskich pochodzą z opracowania Human Development Report Office z 1995 roku. Za jednostkę przyjęto tysiąc dolarów, dzieląc oryginalne wartości przez 1000.

Zbiór danych ograniczono do prób ogólnokrajowych pochodzących z Europy. W związku z tym analizy objęły 16 krajów: Niemcy, Wielką Brytanię, Austrię, Węgry, Irlandię, Holandię, Norwegię, Szwecję, Czechy, Bułgarię, Słowenię, Polskę, Rosję, Hiszpanię, Łotwę i Słowację.

Tabela 1 przedstawia kolejne fazy modelowania hierarchicznego. Pierwszy model, nazywany zerowym, nie zawiera żadnego predyktora. Dokonuje się w nim dekompozycji wariancji zmiennej zależnej na wewnątrz- i międzygrupową. Równania regresji w poszczególnych grupach pozbawione są współczynników nachylenia, mają jedynie wartości stałej. Oznacza to, że w modelu zerowym wynik osoby badanej można przewidywać jedynie w oparciu o średnią wartość zmiennej zależnej w grupie, do której osoba ta należy.

Dekompozycja wariancji pozwala za pomocą współczynnika korelacji wewnątrzklasowej (ICC) oszacować stopień podobieństwa odpowiedzi w badanych krajach. Dzięki ICC badacze mogli stwierdzić, w jakim stopniu indywidualny poziom ksenofobii zależy od kraju pochodzenia respondenta. Wartości umieszczone w Tabeli 1 wskazują, że ICC wyniósł 0,17 (na podstawie formuły $\tau_{00}/\tau_{00} + \sigma^2 = 0{,}127/0{,}127 + 0{,}627$), co pokazuje, że 17% międzyosobniczej wariancji ksenofobii generują tylko i wyłącznie różnice związane z krajem pochodzenia respondenta.

Kolejnym etapem hierarchicznej analizy danych jest model 1, przedstawiający zmiany w składowych wariancji po wprowadzeniu predyktora z poziomu 1 (w omawianym przykładzie jest nim tożsamość narodowa). Wyniki w Tabeli 1 wskazują, że respondenci przejawiający inkluzywny (otwarty) wzorzec tożsamości narodowej są mniej ksenofobiczni niż osoby z wzorcem ekskluzywnym (wartość estymatora efektu, oznaczana jako Val, wynosi –0,418 i jest istotna statystycznie na poziomie $p < 0{,}01$). Wprowadzenie predyktora zmiennej zależnej skutkuje redukcją wariancji wewnątrzgrupowej – porównanie wartości σ^2 wskazuje na jej zauważalny spadek (0,574 < 0,627). Na podstawie różnic między estymatorami σ^2 w modelach 0 i 1 możemy powiedzieć, że predykator z poziomu 1 wyjaśnia 8,5% wariancji ksenofobii [1 – (0,574/0,627) · 100%]. Okazało się też, że predyktor z poziomu 1 ma pewien wpływ na wariancję grupowych stałych. Spadła ona z 0,127 do 0,115, dzięki czemu możemy powiedzieć, że tożsamość narodowa tłumaczy 9,5% międzygrupowej zmienności ksenofobii. Badane kraje różniły się średnim nasileniem ksenofobii, a różnice te wiązały się systematycznie z poziomem inkluzywności tożsamości narodowej [1 – (0,115/0,127) · 100% = 9,5%].

O tym, czy poprawa dopasowania modelu do danych osiągnięta w modelu 1 jest istotna statystycznie, informuje test różnicy logarytmicznego wskaźnika wiarygodności. Wskaźnik ten, przyjmujący wartości ujemne o rozkładzie statystyki $\chi^2_{df = (Dt);(Db)}$, opisuje tzw. dobroć dopasowania modelu na podstawie oceny, na ile prawdopodobne jest uzyskanie empirycznych wyników w danej próbie przy określonych wartościach estymatorów parametrów w populacji. Konwencjonalnie oznaczany jest symbolem L (*likelihood*) a jego logarytm (LL) obliczamy z wzoru LL = 2 · [ln(D_b) – ln(D_t)], gdzie D_b oznacza wskaźnik wiarygodności w modelu bazowym, D_t wskaźnik wiarygodności w modelu testowanym, *df*(D_b) liczbę stopni swobody w modelu bazowym, a *df*(D_t) liczbę stopni swobody w modelu testowanym. Model jest doskonale dopasowany do danych, gdy wartość LL = 0.

Wyniki przedstawione w Tabeli 2 wskazują, że w modelu 1 wartość logarytmu wskaźnika wiarygodności jest wyższa niż w modelu zerowym (–22467,85 > –23343,43). Oznacza to redukcję początkowej wielkości niedopasowania, czyli – innymi słowy – lepsze dopasowanie do danych empirycznych modelu 1. Ponadto różnica w stopniu dopasowania wynosząca 1751,15 (2 · 875,6, czyli podwojona wartość różnicy między modelami 0 i 1) jest przy 1 stopniu swobody (4 – 3 = 1) istotna statystycznie na poziomie $\alpha = 0{,}01$.

Model 1 kończy etap wyjaśniania wariancji na poziomie indywidualnym (σ^2). Od tej chwili przedmiotem zainteresowania badacza będą składowe wariancji istniejącej na poziomie grupowym – τ_{00} i τ_{11}. Jako predyktor z tego poziomu w modelu 2 włączono do analiz wielkość PKB *per capita*. Zgodnie z zasadami działania modeli hierarchicznych zmienne z poziomu makro wyjaśniają wariancję międzygrupową (wartości stałych) τ_{00}, która w tym przypadku jest efektem różnic średniego nasilenia ksenofobii w poszczególnych krajach.

Wyniki analizy regresji dla modelu 2 znajdziemy w Tabeli 1. Okazuje się, że efekt PKB *per capita* jest istotny statystycznie i ma kierunek zgodny z oczekiwaniami (Val = – 0,021; $p < 0{,}01$). Potwierdziła się więc hipoteza, zgodnie z którą im niższy PKB na głowę mieszkańca, tym wyższy kolektywny (grupowy) poziom ksenofobii. Test istotności różnicy dopasowania, przedstawiony w Tabeli 2, wskazuje, że w modelu 2 następuje kolejna redukcja niedopasowania, czyli wzrost logarytmu wskaźnika wiarygodności (–22463,25 > –22467,85), a wielkość różnicy LL, wynosząca 9,19, jest przy jednym stopniu swobody istotna na poziomie $\alpha = 0{,}01$. Zmienna PKB *per capita* wyjaśnia 44% wariancji międzygrupowej (między krajami). Odsetek ten otrzymujemy po podstawieniu odpowiednich wartości z modelu 1 i 2 [1 – (0,064/0,115) · 100% = 44,3%].

Konkluzja, iż model 2 jest lepiej (niż model 1) dopasowany do danych empirycznych zamyka etap modelowania, w którym skupialiśmy się na wyjaśnieniu wariancji międzygrupowej τ_{00}. Modele 3 i 4 służą testowaniu hipotezy badawczej, według której negatywny związek inkluzywnej tożsamości narodowej i ksenofobii będzie najsil-

Tabela 1.
Hierarchiczna analiza regresji w przykładzie z badaniem międzykulturowym

Efekty \ Model	Model 0			Model 1			Model 2			Model 3			Model 4		
	Val	SE	p	Val	SE	p	Val	SE	p	Val	SE	p	Val	SE	p
EFEKTY STAŁE															
Poziom indywidualny															
Stała (Υ_{00})	3,242	0,089	< 0,01	1,935	0,090	< 0,01	2,317	0,129	< 0,01	2,377	0,193	< 0,01	2,756	0,301	< 0,01
Tożsamość inkluzywna (Υ_{10})				–0,418	0,010	< 0,01	–0,418	0,010	< 0,01	–0,414	0,053	< 0,01	–0,279	0,098	< 0,01
Poziom grupowy															
PKB *per capita* (Υ_{01})							–0,021	0,006	< 0,01	–0,023	0,006	< 0,01	–0,043	0,014	0,01
PKB *per capita* · tożsamość (Υ_{11})													0,037	0,009	< 0,01
EFEKTY LOSOWE															
σ^2	0,627	0,792		0,574	0,758		0,574	0,758		0,561	0,749		0,561	0,749	
τ_{00}	0,127	0,357		0,115	0,338		0,064	0,254		0,403	0,635		0,353	0,594	
τ_{11}										0,044	0,209		0,021	0,193	

Tabela 2.
Test istotności statystycznej różnic w dopasowaniu kolejnych modeli

Model	*df*	LL	Test	Różnica LL	*p*
0	3	–23343,43			
1	4	–22467,85	0 *vs.* 1	1751,15	< 0,01
2	5	–22463,25	1 *vs.* 2	9,19	< 0,01
3	7	–22267,61	2 *vs.* 3	391,28	< 0,01
4	8	–22210,43	3 *vs.* 4	114,33	< 0,01

LL – logarytmiczny wskaźnik wiarygodności niejszy w krajach cieszących się najwyższym poziomem PKB *per capita* i najsłabszy w krajach najbiedniejszych. Zatem kolejny etap modelowania dotyczyć będzie wariancji współczynników B, reprezentującej efekty interakcyjne.

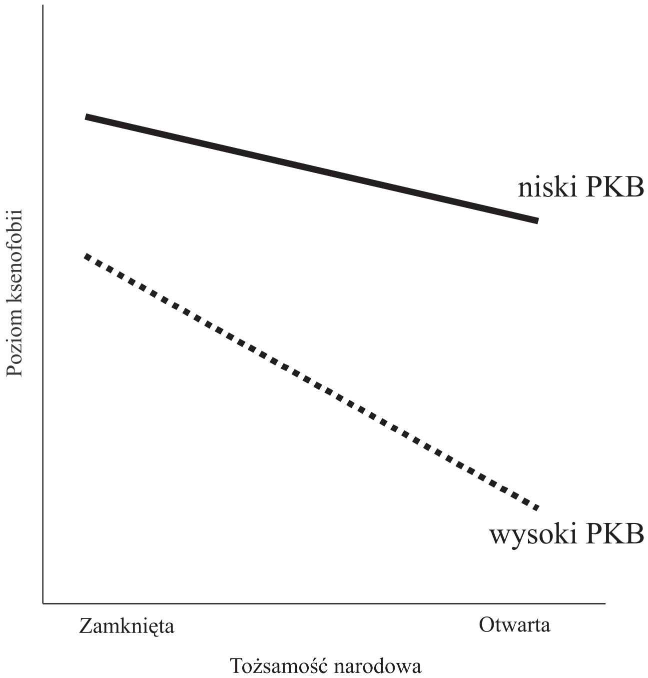
Aby sprawdzić hipotezę o interakcji, w modelu 3 włączono do analizy hierarchicznej efekt losowy predyktora z poziomu 1 (tożsamość narodowa), co pozwoliło na oszacowanie parametrów nachylenia linii regresji (współczynnik B) oddzielnie dla każdego kraju i oszacowanie wariancji grupowych współczynników B (τ_{11}). Z Tabeli 2 dowiadujemy się, że model zezwalający na wariancję współczynników nachylenia (model 3) jest lepiej dopasowany do danych empirycznych niż poprzedzający go model 2 (–22267,61 < –22463,25), a różnica w wielkości LL wynosi 391,28 i przy dwóch stopniach swobody (7 – 5 = 2) jest istotna statystycznie na poziomie $\alpha = 0{,}01$. Wnioskujemy z tego, że między grupami (krajami) istnieją znaczące różnice w sile związku inkluzywnej tożsamości narodowej i ksenofobii. Oszacowana w modelu 3 wariancja współczynników nachylenia (τ_{11}) wynosi 0,044.

Ponieważ poszukujemy czynnika mogącego wyjaśniać wariancję parametrów nachylenia regresji oszacowanych odrębnie dla każdego kraju, musi on być immanentną właściwością krajów (czy raczej społeczeństw), a nie osób badanych. W tej sytuacji, zgodnie z hipotezą trzecią sprawdzamy, czy moderatorem tym może być PKB *per capita* (predyktor poziomu 2). Warto zauważyć, że jest to schemat postępowania bardzo zbliżony do testowania efektu interakcji w klasycznej analizie regresji. Jedyna różnica polega na tym, że w analizie hierarchicznej mamy do czynienia z interakcją międzypoziomową (zmienna z poziomu grupowego wywiera wpływ na zmienne mierzone na poziomie indywidualnym).

Z Tabeli 1 dowiadujemy się, że efekt interakcji uwzględniony w modelu 4 jest istotny statystycznie (Val = 0,037; $p < 0{,}01$). Wpływ PKB *per capita* wyjaśnia 52,3% całkowitej wariancji współczynników B, oszacowanej w modelu 3 [1 – (0,021/0,044) · 100%]. Dodatkowo, model 4 jest lepiej dopasowany do danych empirycznych niż model 3 (–22210,43 > –22267,61), a różnica między nimi – wynosząca 114,36, jest przy jednym stopniu swobody (8 – 7 = 1) istotna statystycznie na poziomie $\alpha = 0{,}01$ (por. Tabela 2).

Graficzny obraz otrzymanych związków pokazuje Rysunek 2, potwierdzając słuszność wszystkich trzech hipotez sformułowanych przez badaczy. Po pierwsze, inkluzywna tożsamość narodowa jest rzeczywiście negatywnym predyktorem ksenofobii; po drugie, ksenofobia jest silniejsza w krajach słabiej rozwiniętych (niski PKB *per capita*); i po trzecie, siła negatywnego związku między inkluzywną tożsamością narodową i ksenofobią jest największa w krajach najlepiej rozwiniętych ekonomicznie (wysoki PKB *per capita*).

Powyższą konkluzję warto uzupełnić o porównanie „mocy wyjaśniającej" odkrytych zależności. Prześledziwszy kolejne etapy budowy modelu (Tabela 1), stwierdzamy, że badaczom udało się wyjaśnić – kolejno, zgodnie z numeracją hipotez – 8,5% wewnątrzgrupowej wariancji zmiennej zależnej (σ^2); 50% (łączny efekt predyktorów z poziomu 1 i 2) wariancji średnich zmiennej zależnej między grupami (τ_{00}); oraz 52,3% międzygrupowej wariancji siły związku (współczynników B) między predyktorem z poziomu 1 i zmienną zależną (τ_{11}).



Rysunek 2.
Efekty proste tożsamości (poziom 1) w zależności od poziomu PKB (poziom 2).

Przykład 2 – badanie eksperymentalne

Przykład ten zaczerpnęliśmy z publikacji autorstwa Cohena i współpracowników (2003). Badanie dotyczyło efektów specjalnego programu odchudzania, w którym wzięło udział 386 kobiet w 40 grupach ćwiczeniowych. Grupy zostały losowo przyporządkowane do jednego z poziomów eksperymentalnych – 24 grupy (230 osób) do warunku z manipulacją eksperymentalną i 16 grup (156 osób) do warunku kontrolnego. W warunku z manipulacją uczestniczki realizowały złożony program odchudzania zawierający dokładną specyfikację diety, doradztwo, ćwiczenia, lekcje przygotowywania dietetycznych posiłków i cotygodniowe spotkaniach, na których opowiadano o swoich sukcesach i ewentualnych porażkach; w warunku kontrolnym odbywały się jedynie cotygodniowe spotkania, podczas których odchudzające się kobiety dzieliły się swoimi doświadczeniami. Przed rozpoczęciem programu wszystkim uczestniczkom zmierzono poziom motywacji do odchudzania.

Obydwa predyktory, które znalazły się w równaniu, zostały wycentrowane. Centrowanie polega na odjęciu wyników surowych od średniej ogólnej – po takim przekształceniu średnia ogólna zmiennej wynosi 0. Dla manipulacji efekt centrowania został osiągnięty poprzez zważenie kodów grupowych. W rezultacie tego zabiegu grupie eksperymentalnej odpowiadała waga 0,404 ([(+1) ·156/(156 + 230)]), a grupie kontrolnej waga –0,596 ([(–1) · · 230/(156 + 230)]).

Równanie mieszanego modelu regresji dla efektów losowych wyglądało następująco:

$$y_{ij} = \Upsilon_{01}\,\text{MANIPULACJA\_C}_j + \Upsilon_{10}\,\text{MOTYWACJA\_C}_{ij} + \\ +\Upsilon_{11}\,\text{MANIPULACJA\_C}_j \times \text{MOTYWACJA\_C}_{ij} + \\ + \Upsilon_{00} + + (u_{0j} + u_{1j}\,x_{ij} + r_{ij})$$

W przykładzie 2 hierarchiczna analiza regresji odpowiada na trzy pytania: 1) czy prawdziwa jest hipoteza, że manipulacja eksperymentalna ma pozytywny wpływ na spadek wagi ciała; 2) czy prawdą jest, że różnice indywidualne w poziomie motywacji są predyktorem spadku wagi ciała; 3) czy istnieje interakcja pomiędzy predyktorem z poziomu 2 (manipulacja) i predyktorem z poziomu 1 (motywacja). Hipoteza badawcza dotycząca interakcji postuluje, że działanie programu będzie zwiększało siłę związku pomiędzy motywacją i utratą wagi. Innymi słowy, manipulacja eksperymentalna powinna odegrać rolę moderatora.

Wyniki modelowania hierarchicznego przedstawia Tabela 2. Podobnie jak w przykładzie 1 rozpoczyna je model zerowy. W tej postaci jest on tożsamy z jednoczynnikową ANOVĄ z efektami losowymi – 40 poziomów czynnika (40 grup) stanowi losową, reprezentatywną próbę wszystkich potencjalnie dostępnych grup. Na tym etapie estymowane są dwie składowe wariancji – τ_{00}, opisująca zróżnicowanie stałych w 40 grupach (międzygrupowa wariancja średniego grupowego spadku wagi ciała), oraz σ^2, czyli wartości reszt dla poziomu 1. Z Tabeli 2 dowiadujemy się, że ich wielkość wynosi, odpowiednio, 4,906 i 16,069. Wartości testu *Z* wskazują, że obydwa estymatory są istotne statystycznie ($Z = 3{,}14$; $p = 0{,}002$ dla τ_{00} i $Z = 13{,}12$; $p = 0{,}001$ dla σ^2). Fakt, że τ_{00} różni się istotnie statystycznie od zera jest dowodem na istnienie znaczącej wariancji średnich grupowych, a tym samym na obecność w zbiorze danych zjawiska „klasteringu". Potwierdzeniem tego jest, obliczona z τ_{00} i σ^2, korelacja wewnątrzklasowa ICC, której wielkość wyniosła 0,24 [4,906/(4,906 + 16,069)]. W kolejnych modelach τ_{00} i σ^2 posłużą do śledzenia efektów wpływu predyktorów z poziomów 1 i 2 na zmienną zależną.

Model 1 wprowadza efekt predyktora z poziomu 1. Dowiadujemy się z niego, do jakiego stopnia poziom motywacji do odchudzania pozwala przewidywać spadek wagi ciała. Po dodaniu predykatora, w losowej części modelu 1 pojawiają się dwie nowe składowe wariancji – pierwsza z nich (τ_{11}) jest efektem międzygrupowej zmienności współczynnika nachylenia; druga (τ_{01}) – uwzględnia możliwość kowariancji (skorelowania) grupowych stałych i współczynników B. Ponieważ wariancja współczynników B jest istotna statystycznie, wiemy już, że siła związku pomiędzy motywacją i spadkiem wagi zmienia się w zależności od grupy ($\tau_{00} = 0{,}933$; $Z = 2{,}48$ dla $p = 0{,}013$). Zatem „klastering" wpływa nie tylko na średnie grupowe, ale także na nachylenie grupowych linii regresji. Dodatnia wartość τ_{01} wskazuje, że im większy grupowy współczynnik przecięcia (średni spadek wagi), tym silniejszy związek pomiędzy predyktorem z poziomu 1 i zmienną zależną. Efekt ten nie jest jednak istotny statystycznie ($\tau_{01} = 0{,}585$; $Z = 1{,}52$ dla $p = 0{,}128$).

Wprowadzenie predyktora z poziomu 1 spowodowało znaczącą redukcję składowych wariancji w stosunku do modelu zerowego. Wielkość σ^2 spadła z 16,069 do 5,933. Oznacza to, że aż 63% wewnątrzgrupowej wariancji zmiennej zależnej (spadek wagi ciała w funtach) tłumaczy siła motywacji. Ponadto okazało się, że predyktor z poziomu 1 ma równie duży wpływ na wariancję grupowych stałych. Spadła ona z 4,906 do 2,397, dzięki czemu możemy powiedzieć, że motywacja tłumaczy 51% międzygrupowej zmienności przeciętnej wielkości spadku wagi ciała. Tak więc, badane grupy różniły się średnią wielkością utraty wagi ciała, a różnice te wiązały się systematycznie z poziomem motywacji w poszczególnych grupach.

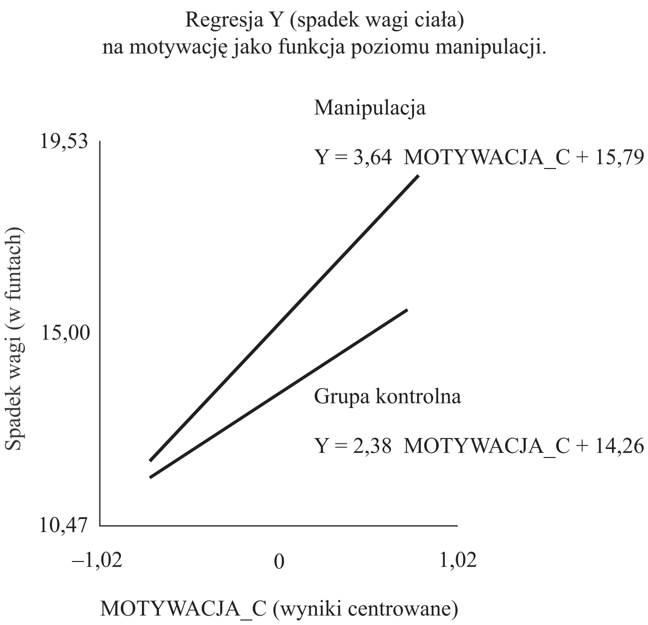
Końcowym elementem hierarchicznej struktury jest model 2. Pokazuje on interakcję manipulacji eksperymentalnej i motywacji, w postaci dodatniego efektu synergii predyktorów z poziomu indywidualnego i grupowego. Poziom 2 (manipulacja) modyfikuje losowe efekty współczynników nachylenia, co oznacza, że siła związku predykatora z poziomu 1 (motywacja) i zmiennej zależnej (spadek wagi ciała) zmienia się w zależności od warunku eksperymentalnego.

Rysunek 3 pokazuje efekty proste regresji zmiennej Y (spadek wagi ciała) na motywację. Wynika z niego, że w całej badanej zbiorowości (386 osób) manipulacja eksperymentalna wywołała zmianę przeciętnego spadku wagi

Tabela 3.
Hierarchiczna analiza regresji w eksperymencie z programem odchudzania

	Estymatory parametrów dla efektów losowych				
	Czynnik grupujący	Estymator	Błąd standard.	*Z*	*p*
Model 0					
τ_{00} – wariancja stałych	GRUPA	4,906	1,560	3,14	0,002
σ^2 – reszty poziomu 1		16,069	1,225	13,12	0,001
Model 1					
τ_{00} – wariancja stałych	GRUPA	2,397	0,741	3,23	0,001
τ_{01} – kowariancja stałych i współczynników B	GRUPA	0,585	0,385	1,52	0,128
τ_{11} – wariancja współczynników B	GRUPA	0,933	0,376	2,48	0,013
σ^2 – reszty poziomu 1		5,933	0,476	12,47	0,001
Model 2					
τ_{00} – wariancja stałych	GRUPA	1,967	0,657	2,99	0,003
τ_{01} – kowariancja stałych i współczynników B	GRUPA	0,145	0,314	0,46	0,645
τ_{11} – wariancja współczynników B	GRUPA	0,556	0,301	1,85	0,065
σ^2 – reszty poziomu 1		5,933	0,475	12,48	0,001

	Estymatory parametrów dla efektów stałych				
	Estymator	Błąd standardowy	*df*	*t*	*p*
STAŁA Υ_{00}	15,166	0,259	38	58,49	0,001
MANIPULACJA_C Υ_{01}	1,528	0,529	38	2,89	0,006
MOTYWACJA_C Υ_{10}	3,130	0,185	344	16,95	0,001
INTERAKCJA Υ_{11}	1,245	0,377	344	3,30	0,001



Rysunek 3.
Efekty proste motywacji (poziom 1) w zależności od manipulacji eksperymentalnej (poziom 2).

ciała z 14,26 do 15,79 funta. Co więcej, spowodowała również wzrost siły związku pomiędzy motywacją i utratą wagi. Jak pokazują równania efektów prostych, o ile w grupie kontrolnej wzrost motywacji o jedną jednostkę pomiaru pozwala przewidywać spadek wagi o 2,38 funta, to w grupie eksperymentalnej przewidywany spadek wynosi 3,64 funta.

W modelach hierarchicznych zmienność współczynników przecięcia i nachylenia powinny wyjaśniać predyktory wprowadzone na poziomie 2. Wielkości estymatorów w modelu 2 (zob. Tabela 3) pokazują wyraźną redukcję składowych wariancji. Wprowadzenie manipulacji eksperymentalnej i interakcji międzypoziomowej spowodowało wyraźny, o 18% (z 2,397 do 1,967), spadek wariancji grupowych stałych (τ_{00}) oraz jeszcze większy, o 35% (z 0,933 do 0,556), spadek wariancji współczynników B (τ_{11}). W obu przypadkach wartości testu *Z* wskazują, że pewna część wariancji pozostaje nadal niewyjaśniona. Jednak tylko dla wariancji współczynników przecięcia wielkość ta jest istotna statystycznie (odpowiednio $Z = 2{,}99$; $p = 0{,}003$ dla τ_{00} i $Z = 1{,}85$; $p = 0{,}065$ dla τ_{11}).

W Tabeli 3, oprócz efektów losowych, znajdujemy również oszacowania efektów stałych w postaci estymatorów czterech parametrów: wartości stałej (Υ_{00}), efektu predyktora z poziomu 2 (Υ_{01}), efektu predyktora z poziomu 1 (Υ_{10}) oraz efektu interakcji międzypoziomowej (Υ_{11}). W modelu efektów stałych otrzymujemy następujące równanie regresji:

$$Y = 1{,}53\ \text{MANIPULACJA\_C} + 3{,}13\ \text{MOTYWACJA\_C} + 1{,}25\ \text{INTERAKCJA} + 15{,}17$$

Tabela 4 przedstawia tradycyjną analizę regresji wykonaną metodą najmniejszych kwadratów na danych zdezagregowanych (ignorującą grupową strukturę danych). Porównanie estymatorów efektów stałych otrzymanych w modelu hierarchicznym (Tabela 3) z wartościami efektów stałych w analizie klasycznej (Tabela 4) wskazuje, że ich wielkości są podobne. Zdecydowanie większe różnice występują w błędach standardowych odpowiadających poszczególnym efektom. Najbardziej rzuca się w oczy różnica towarzysząca manipulacji eksperymentalnej – w modelu hierarchicznym błąd standardowy estymatora wynosi 0,529, a w modelu klasycznym tylko 0,301.

Model hierarchiczny traktuje zmienną eksperymentalną w taki sposób, jakby jej rozkład dotyczył 40 obserwacji (po jednej dla każdej grupy); w modelu klasycznym rozkład zmiennej eksperymentalnej obejmuje 386 pojedynczych obserwacji. Odmienny status zmiennej skutkuje w tradycyjnym modelu regresji niedoszacowaniem błędu standardowego estymatora. A to, jak już wcześniej podkreślaliśmy, prowadzi do zjawiska inflacji poziomu alfa, czyli znaczącego wzrostu prawdopodobieństwa popełnienia błędu pierwszego rodzaju.

Uwagi końcowe

Artykuł ten – z oczywistych względów – jest jedynie krótkim wprowadzeniem do problematyki hierarchicznych modeli liniowych. Opisaliśmy w nim podstawowe metodologiczne przesłanki zastosowania tej metody, akcentując jej zalety w porównaniu z klasyczną analizą regresji opartą na metodzie najmniejszych kwadratów; omówiliśmy szczegółowo konceptualną postać modelu – podział na efekty stałe i losowe, wielopoziomową strukturę (z uwzględnieniem interakcji międzypoziomowej), specyficzne ujęcie składowych wariancji; na koniec przedstawiliśmy dwa przykłady empirycznej aplikacji samej metody, poparte szczegółową interpretacją wyników.

Modelowanie hierarchiczne to metoda złożona i stosunkowo nowa. Świadomie pominęliśmy tutaj lub ograniczyliśmy do koniecznego minimum wiele zagadnień tyleż istotnych, co szczegółowych. Dotyczą one m.in. testowania istotności statystycznej efektów stałych i losowych (w tym drugim przypadku istnieje kilka alternatywnych metod), różnic związanych z mocą testów statystycznych na poziomie mikro i makro, problemu estymacji równania regresji dla efektów losowych w poszczególnych grupach, centrowania danych itd. Czytelnika chcącego lepiej zgłębić wymienione problemy odsyłamy do powszechnie polecanych opracowań oferujących szczegółowe wprowadzenie do hierarchicznej analizy danych (Bryk i Raudenbush, 1992; Goldstein, 1995; Hox, 2002; Kreft i de Leeuw, 1998; Snijders i Bosker, 1999). O wiele obszerniejsze niż ten artykuł, przystępnie napisane i ilustrowane przykładami wprowadzenie oferują poświęcone tej metodzie, najpopularniejsze programy komputerowe – MlwiN (Rasbash, Steele, Browne i Goldstein, 2009) i HLM (Raudenbush, Cheong i Bryk, 2004). Bezpłatne, demonstracyjne wersje tych programów można pobrać ze stron internetowych.

Na zakończenie chcemy jeszcze wspomnieć o dwóch bardzo ważnych kwestiach. Zacznijmy od wspominanego kilkakrotnie problemu inflacji poziomu alfa. Czytelnik powinien wiedzieć, że występujące pomiędzy oboma modelami (klasycznym i hierarchicznym) różnice w błędach standardowych estymatorów, wynikają nie tylko z faktu, że hierarchiczne modele efektów losowych – w przeciwieństwie do klasycznej analizy regresji – poddają statystycznej kontroli wariancję związaną z „klasteringiem” w zbiorze danych. Drugim źródłem rozbieżności są odmienne procedury estymacji parametrów modelu. Zauważyliśmy już wcześniej, że w klasycznym modelu

Tabela 4.
Analiza regresji dla efektów stałych (metoda najmniejszych kwadratów) w eksperymencie z programem odchudzania

	Estymator	Błąd standardowy	*df*	*t*	*p*
STAŁA	15,105	0,148			
MANIPULACJA_C	1,578	0,301	1	5,239	< 0,001
MOTYWACJA_C	3,330	0,145	1	22,968	< 0,001
INTERAKCJA	1,446	0,300	1	4,820	< 0,001

efektów stałych estymatory współczynników w populacji otrzymujemy dzięki metodzie zwykłych najmniejszych kwadratów, a w modelu efektów losowych za pomocą metody największej wiarygodności (ML) lub blisko z nią powiązanej metody ograniczonej największej wiarygodności (REML). Parametry wyestymowane przy użyciu procedury ML lub REML różnią się często od otrzymanych dzięki procedurze najmniejszych kwadratów, ponieważ wieloetapowy proces iteracji, składający się często z setek kroków, ma na celu znalezienie w badanej próbie takiego zbioru estymatorów, które są najbardziej wiarygodnym przybliżeniem parametrów w populacji.

I druga ważna uwaga. Przypuszczamy, że jedno z najczęściej pojawiających się pod wpływem tego tekstu pytań będzie brzmiało: kiedy powinniśmy stosować model hierarchicznej analizy regresji dla efektów losowych, a kiedy klasyczny model analizy regresji dla efektów stałych z wykorzystaniem zmiennych instrumentalnych? Obydwa podejścia wydają się ekwiwalentne, ponieważ pozwalają na statystyczną kontrolę zgrupowanej, wielopoziomowej struktury danych. Do wyboru metody hierarchicznej z pewnością skłania chęć uniknięcia technicznych uciążliwości związanych – zwłaszcza w modelach wielozmiennowych – z kodowaniem zmiennych instrumentalnych (dla efektów głównych i interakcyjnych). Inne względy mogą jednak przemawiać za użyciem metody tradycyjnej.

Wskazówek i rekomendacji dotyczących wyboru jednej z dwóch metod udzielają Snijders i Bosker (1999), argumentując, że decyzja powinna zależeć od następujących czynników: liczby grup, liczebności poszczególnych grup, rozkładów reszt regresji dla predyktorów z obu poziomów, procedury doboru próby i zakresu generalizacji wyników. Zalecają oni podejście klasyczne (kodowanie zmiennych instrumentalnych), kiedy liczba grup jest stosunkowo niewielka (mniej niż 10) oraz gdy poszczególne grupy wykazują bardzo specyficzne właściwości (np. kategoria wykształcenia, wyznanie religijne, grupa etniczna), a badacz chciałby pokazać tę specyfikę w ujęciu porównawczym. Podejście hierarchiczne jest zalecane, gdy badane grupy należy traktować jedynie jako próbę losową z populacji grup, a celem badacza jest generalizacja wyników na całą zbiorowość (np. z próby kilkudziesięciu rodzin na pełną populację rodzin). Również niewielkie liczebności obserwacji w grupach skłaniają do wyboru metody hierarchicznej, ponieważ w tym przypadku estymatory parametrów w poszczególnych grupach posiłkują się rozkładem wyników w całym zbiorze danych.

LITERATURA CYTOWANA

Berry, J. W., Poortinga, Y. H., Segall, M. H., Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications.* Cambridge: University Press.

Bryk, A. S., Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Coenders, M., Lubbers, M., Scheepers, P. (2007). From a distance. Avoidance of social contacts with immigrants in the European Union. W: E. Poppe i M. Verkuyten (red.), *Culture and conflct* (s. 217–243). Amsterdam: Aksant Academic Publisher.

Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum.

Goldstein, H. (1995). *Multilevel statistical models*. London: Arnold.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT–Publikaties.

Hox, J. J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Erlbaum.

Kreft, I., de Leeuw, J. (1998). *Introducing multilevel modeling.* London: Sage Publications.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. (2009). *Nlme: Linear and nonlinear mixed effects models. R package version 3.1-96.* Vienna: R Foundation for Statistical Computing.

R Development Core Team (2009). *R: A language and environment for statistical computin.* Vienna: R Foundation for Statistical Computing.

Rasbash, J., Steele, F., Browne, W. J., Goldstein, H. (2009). *A user's guide to MlwiN 2.10.* Bristol: Centre for Multilevel Modelling, University of Bristol.

Raudenbush, S. W., Cheong, Y. F., Bryk, A. S. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International, Inc.

Snijders, T., Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

West, S. G., Aiken, L. S., Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality*, *64,* 1–48.

PRZYPISY

1. Zapis matematyczny w modelach hierarchicznych jest dość rozbudowany. W tekście zastosowano notację powszechnie przyjętą w literaturze przedmiotu (Raudenbush i Bryk, 2002: za Cohen i in., 2003). W pełnej postaci została ona umieszczona w aneksie na końcu tekstu. W obrębie pojedynczej analizy regresji z efektami losowymi mamy g grup. Używamy indeksu i i j dla oznaczenia dowolnej obserwacji i w dowolnej grupie j.

2. W modelowaniu hierarchicznym może pojawić się więcej niż jeden subpoziom poziomu makro. Na przykład klasy szkolne w obrębie szkół (poziom nadrzędny); szkoły w obrębie dzielnic, dzielnice w obrębie miast itd.

3. Istnieje wiele innych aplikacji, za pomocą których można wykonywać analizy hierarchiczne. Najbardziej znane specjalistyczne programy to HLM (Raudenbush i in., 2004) i MlwiN (Rasbash i in., 2009).

PODSTAWOWE TERMINY

Zapis modelu analizy regresji dla efektów losowych w podejściu hierarchicznym

A. Współczynniki dla równania regresji w grupie (poziom mikro)
   B_{0j} = współczynnik przecięcia (stała) w grupie j na poziomie 1
   B_{1j} = współczynnik nachylenia w grupie j na poziomie 1

B. Współczynniki regresji w populacji: efekty stałe w modelu
   Υ_{00} = współczynnik przecięcia w populacji
   Υ_{10} = współczynnik przecięcia w populacji dla regresji zmiennej zależnej Y na predykator 1 poziomu
   Υ_{01} = współczynnik nachylenia w populacji dla regresji zmiennej zależnej Y na predykator 1 poziomu
   Υ_{11} = współczynnik regresji w populacji dla interakcji pomiędzy predykatorami z poziomu 1 i 2

C. Reszty i składowe wariancji: efekty losowe w modelu
   1. Reszty
      r_{ij} = błąd dla obserwacji i w grupie j (równanie dla poziomu 1)
      u_{0j} = losowy efekt odchylenia współczynnika przecięcia w grupie j od współczynnika przecięcia w populacji (równanie dla poziomu 1)
      u_{1j} = losowy efekt odchylenia współczynnika nachylenia w grupie j od współczynnika przecięcia w populacji (równanie dla poziomu 1)
   2. Składowe wariancji
      σ^2 = losowa wariancja błędu na poziomie mikro (wariancja r_{ij})
      τ_{00} = losowa wariancja współczynników przecięcia (wariancja u_{0j})
      τ_{11} = losowa wariancja współczynników nachylenia (wariancja u_{1j})
      τ_{01} = kowariancja w modelu analizy regresji z efektami losowymi (kowariancja pomiędzy u_{0j} i u_{1j})

Hierarchical linear models. On their advantages and reasons for application

Piotr Radkiewicz[1], Marcin W. Zieliński[2]

[1] *Institute of Psychology, Polish Academy of Sciences*
[2] *Institute for Social Studies, University of Warsaw*

Abstract

Independence of observations is one of the key assumptions underlying regression analysis and other methods based on the general linear model. The assumption of independence of observations is met, when a score on an outcome variable obtained by an individual is not dependent on results of other persons. This article introduces the hierarchical linear modeling (HLM) – statistical method that is recommended, when there is a real chance, that the assumption of observations' independence is violated. The structure of our article is threefold. In the first part we present basic methodological reasons for applying HLM method, stressing its advantages in comparison to the traditional regression analysis based on the ordinary least squares estimation. The second part introduces the most important theoretical notions underlying hierarchical models – a division into fixed and random effects, a multilevel data structure (including cross-level interaction), and a specific approach to variance components. In the third part we show two empirical examples of HLM application, including a detailed interpretation of their results.

Key words: hierarchical linear models, independence of observations, intraclass correlation, random effects model, variance components

Złożono: 28.05.2010
Złożono poprawiony tekst: 3.10.2010
Zaakceptowano do druku: 3.10.2010