

Henryk Domański  
Polska Akademia Nauk

### **Prawidłowości społeczne w tablicach**

**Zbigniew Sawiński:** *Zastosowania tablic w badaniach zjawisk społecznych*. Warszawa: Wydawnictwo IFiS PAN 2010, 400 s.

Recenzowana książka sytuuje się w dziedzinie metodologii, a dokładniej – w tym jej aspekcie, który dotyczy analizy danych. Jest monograficznym studium pewnego podejścia w badaniu zjawisk społecznych, gdy ujmowane są one w postaci tablic liczbowych. W naukach społecznych tablice są naturalnym, a czasami jedynym narzędziem analizy zależności między zmiennymi, głównie ze względu na „bardziej jakościowy” (w odróżnieniu od świata przyrody) charakter stosunków społecznych.

Przedmiotem monografii Sawińskiego jest perspektywa „eksploracyjna” – polegająca na identyfikowaniu prawidłowości takimi, jakie one są – w odróżnieniu od perspektywy „konfirmacyjnej”, której istotą jest testowanie hipotez. Do podejścia eksploracyjnego – rzadziej stosowanego w socjologii ilościowej – „przypisane” są pewne techniki analiz statystycznych, takie jak analiza kanoniczna czy analiza korespondencji. Kwestie lokujące się w tym właśnie obszarze podejmuje w swej pracy Sawiński. Celem jego książki jest wykazanie, że za danymi w tablicach liczbowych kryją się określone prawidłowości, odzwierciedlające logikę mechanizmów społecznych, a następnie: (i) omówienie metod identyfikacji tychże, (ii) prezentacja wybranych narzędzi ich analizy, (iii) wydobycie zachodzących między nimi powiązań formalnych – łączy je odwołanie się do własności modelu niezależności stochastycznej i posługiwanie się rozkładem chi-kwadrat.

Wypunktowałem to, ponieważ brakuje mi w miarę jasnego wyspecyfikowania celu książki przez autora. Sformułowanie ze wstępu, że „poświęcona jest ona praktyce analizy wyników badań” i problemom, „jakie napotykać badacze analizując zawartość tablic” (s. 16) można by odnieść do szerokiej klasy analiz. Jest to jedno z niewielu moich zastrzeżeń pod adresem tej pracy, która jest – w moim przekonaniu – jedną z tych niewielu, napisanych przez polskich socjologów, z których można się czegoś nowego nauczyć.

Problematyka „analizy tablic” zebrana jest w siedmiu rozdziałach. Dwa pierwsze są omówieniem podstawowych pojęć i spraw warsztatowych. Dla empiryka mogą to być rzeczy proste, ale – oceniając z perspektywy pięciu następnych rozdziałów – powiedziałbym, że niezbędne, jeżeli chce się je dobrze zrozumieć. Przykładem może być wpływ sposobu grupowania jednostkowych obserwacji na wyniki analiz i problemy z interpretacją danych wieloodpowiedziowych (gdy badanemu nie stawia się żadnych ograniczeń w liczbie możliwych odpowiedzi), uwzględnianie kategorii

rezydujących i interpretacja danych ważonych. Rozdziały te są wprowadzeniem do bardziej substancyjnych problemów.

Interesujące zagadnienia zaczynają się w rozdziale 3, poświęconym prezentacji modelu niezależności stochastycznej, będącego punktem wyjścia dla większości analiz ujmowanych w tabelach. Czym różni się model niezależności od równości szans, jak ją testować, jak oddzielić komponent „równości szans” od „efektu marginesów” i po co to robić? Z kolei rozdział 4 zawiera omówienie różnych mierników, użytecznych dla identyfikacji wzorów zależności rozpatrywanych w tablicach. Sawiński przedstawia ich właściwości, które są – jak podejrzewam – ogółowi nieznanne (jak np. związku indeksu Quateleta z chi-kwadrat).

W rozdziale 5 czytelnik dowiaduje się, jak można uchwycić dystanse między poszczególnymi kategoriami analizowanych zmiennych, posługując się metodami „różnic między profilami” i „dopasowywania średnich”. Autor wskazuje również na ich ograniczenia, co prowadzi go do omówienia własności modelu kanonicznego w rozdziale 6. Z punktu widzenia analitycznych możliwości podejścia eksploracyjnego, rozdział ten ma znaczenie centralne. Analiza kanoniczna pozwala ustalić siłę związku między analizowanymi zmiennymi, i zdekomponować go na części składowe – umożliwi to identyfikację „zagęszczeń” i „rozrzedzeń” w tablicy zależności, które dają się interpretować w terminach kryjących się za tymi aspektami prawidłowości społecznych. W ostatnim – 7 rozdziale – Sawiński omawia analizę korespondencji i uzasadnia jej przydatność dla identyfikacji wzorów zależności, koncentrując się na aspekcie wizualizacji ustaleń, które zapewnia CA.

Właściwości tych metod znane są specjalistom od analiz ilościowych, a np. analiza korespondencji (CA) znajduje coraz szersze zastosowanie w praktyce badawczej. Nikt jednak przed Sawińskim nie powiązał ich w całość, nie pokazał kryjących się za nimi założeń, możliwości aplikacyjnych i interpretowania uzyskiwanych parametrów (wyników). Cechą prawdziwej nauki jest oryginalność ustaleń – trzeba było przenikliwego spojrzenia, kompetencji i wiedzy autora tej książki, żeby wykazać, że model kanoniczny integruje wiele tych podejść i pozwala (w statystycznym sensie) „wyjaśnić” najistotniejsze aspekty zależności między zmiennymi. Dotyczy to również bardziej szczegółowych kwestii, związanych np. z możliwością dekompozycji zależności (za pomocą mierników odwołujących się do chi-kwadrat) w celu zrozumienia, co np. stoi za wpływem pochodzenia na pozycję zawodową jednostek (klasyczny problem ruchliwości społecznej) lub za wpływem pochodzenia na osiągnięcia edukacyjne (problematyka nierówności w dostępie do wykształcenia).

Nie musi to być punktem odniesienia do oceny tej książki, ale w polskiej socjologii ilościowej zaklasyfikowałbym ją do kategorii pionierskich. Może nie do końca, jako że niektóre z tych problemów zostały zasygnalizowane we wcześniejszej publikacji Sawińskiego (poświęconej analizie maksymalnej korelacji) w „Studiach Socjologicznych” z 1985 roku. Z kolei, w literaturze światowej istnieje wiele opracowań poświęconych omówieniu analizy kanonicznej (z bardziej starych i znanych – np. rozdział w książce Maxwella), przeważnie w połączeniu z innymi technikami statystycznymi z grupy technik taksonomicznych (analiza czynnikowa, dyskryminacyjna itp.), ale koncentrują się one na określonych aspektach (formalnych, oblicze-

niowych), przy pominięciu „genetycznych” związków tego podejścia z właściwościami rozkładów statystycznych, które wydobywa Sawiński.

Z oryginalnością poznawczą łączą się walory dydaktyczne. Tematyczna struktura tej książki ma logikę wykładu. Autor zaczynając od spraw elementarnych przechodzi do bardziej złożonych, omawiając krok po kroku właściwości rozpatrywanych metod statystycznych i konstrukcji indeksów, ilustruje posługiwanie się nimi na konkretnych przykładach, mówi, jak należy je interpretować, a jakich wniosków unikać, i jakie interpretacyjne pułapki związane są np. ze zbyt małą liczebnością niektórych komórek w tablicy dwuwymiarowego rozkładu. Przestudiowanie tej książki powinno ułatwić zrozumienie i praktyczne opanowanie zasad analizy kanonicznej i pokrewnych jej metod nawet mniej analitycznym umysłom (warto między innymi zapoznać się z interpretacją parametru „bezwładności” w analizie korespondencyjnej, która jest przedstawiona w sposób wyjątkowo przystępny).

Z dydaktycznego punktu widzenia można tę książkę traktować jako (chyba pierwszy) systematyczny podręcznik dotyczący technik statystycznych poświęconych analizie struktury zależności między zmiennymi. Ale nie tylko – cztery pierwsze rozdziały poleciłbym wszystkim osobom uczącym się zastosowań statystyki w naukach społecznych, w celu rozszerzenia horyzontów poznawczych, np. odnośnie do kwestii, że mierniki siły związku można podzielić na: (i) opierające się na założeniu o normalności rozkładu (między innymi współczynnik korelacji Pearsona) i (ii) redukcji błędów przewidywania rozkładów – Goodman-Kruskal (rozdział 3.11).

Książka ta może być również podręcznikowym exemplum pisania o rzeczach nietrywialnych w odbiorze. Autor posługuje się językiem przyjaznym dla czytelnika – logicznym, precyzyjnymi zdaniami, z robieniem przerywników (nieszablonowym pomysłem jest zamieszczenie biogramów wybitnych statystyków, takich jak Agresti czy Goodman), a równocześnie napisane jest to z dystansem, dzięki komentarzom odautorskim i ogólniejszej refleksji. Interesujący jest szkic historyczny o rozwoju tych metod, chociaż nie pisze jej historię nauki.

Co poprawiłbym, gdybym był autorem tej książki?

Do dyskusji jest sformułowany na s. 25 pogląd autora, że „w pewnym uproszczeniu” głównym kryterium podziału badań na jakościowe i ilościowe jest to, że zadaniem tych pierwszych jest odtwarzanie mechanizmów, a tych drugich – zasięgu zjawisk społecznych. Co do zasięgu (że nie chwytają go analizy jakościowe) zgoda. Można jednak argumentować, że mechanizmy zjawisk (czyli zdaniem autora „istota”) są również do zidentyfikowania na gruncie metod ilościowych, czego przykładem jest pogłębiona analiza zależności przyczynowych w modelu ścieżkowym. Pisząc o mechanizmach autor ogranicza je, prawdopodobnie, do zjawisk psychospołecznych, a nie uwzględnia zjawisk ze sfery „obiektywnej”, np. mechanizmów stratyfikacji społecznej, ale nieznacznie się w tej kwestii różnimy.

W zakończeniu rozdziału 6 autor – wchodząc na tory ogólniejszej refleksji – przypisuje brak zainteresowania metodami dekompozycji tablic występowaniem między innymi efektu mody (lub braku mody), odwołując się do przykładu mody na modelowanie logarytmiczno-liniowe. Zgadzam się z tymi argumentami, ale tylko częściowo, ponieważ za atrakcyjnością modelowania logarytmiczno-liniowego

kryją się również pewne jego zalety, których Sawiński jest oczywiście świadomy. Może należało by dodać jakieś obszerniejsze wyjaśnienie, że głównym powodem mody na modelowanie logarytmiczno-liniowe jest przekonanie o jego użyteczności w testowaniu hipotez, z kolei za niepopularnością metod dekompozycji kryje się (uproszczone na ogół) przekonanie, że eksploracja jest poruszaniem się po omacku. Mimo że sam posługuję się metodami dekompozycji (zresztą w jakimś zakresie z inspiracji autora tej książki), zaliczyłbym się do badaczy podkreślających zalety modelowania logarytmiczno-liniowego, szczególnie w analizach międzykrajowych. Dostarczają one między innymi w miarę jednoznacznego kryterium podobieństw i różnic między krajami np. w sile ruchliwości społecznej, homogamii małżeńskiej lub wzorów wyboru znajomych, podczas gdy w analizie kanonicznej można to co najwyżej oceniać na oko. Z tych samych powodów metody dekompozycji nie mają też tak dobrego zastosowania do analizy zmian w czasie. Posługując się nimi nie można np. dokładnie określić głębi zmian w zakresie otwartości barier społecznych, chociaż pozwalają one prześledzić kształtowanie się zmian pod względem łatwości zawierania małżeństw między podstawowymi segmentami struktury społecznej, takimi jak inteligencja i chłopi. Wątek o modzie lepiej byłoby rozwijać dopiero po zestawieniu zalet i mankamentów obu tych podejść.

W świetle tylu pochwał, moja krytyka musi być z konieczności zdawkowa. W odniesieniu do polskiej socjologii trudno byłoby mi wskazać bardziej wnikliwą i potrzebną książkę, napisaną przez badacza rozumiejącego strukturę zależności zjawisk społecznych.