

PAWEŁ RYDZEWSKI

METODY ANALIZY DANYCH W SOCJOLOGICZNYCH BADANIACH ROZWODÓW

WSTĘP

Można wyodrębnić kilka typów socjologicznych badań nad rozwodami w zależności od tego, kto lub co stanowi podstawowe źródło informacji:

1. badania terenowe: (a) badania z udziałem osób rozwiedzionych i/lub członków ich rodzin; (b) badania opinii społecznej na temat rozwodów i osób rozwiedzionych;
2. badania dokumentarne: (a) źródłowe – wykorzystujące akta sądowe spraw rozwodowych; (b) statystyczne – wykorzystujące statystyki rozwodowe.

Głównym celem tego artykułu jest udzielenie odpowiedzi na pytanie: Jakie metody analizy danych najczęściej stosuje się w terenowych badaniach rozwodów z udziałem osób rozwiedzionych i/lub członków ich rodzin? Bazę źródłową stanowił wybór socjologicznych artykułów opublikowanych w latach 1980-2000 na łamach renomowanych amerykańskich czasopism naukowych o rozwodach¹. W artykule wykorzystano jedynie niewielką część przeanalizowanych publikacji jako ilustracje omawianych zagadnień.

Dr PAWEŁ RYDZEWSKI – adiunkt Katedry Sociologii WSPA w Lublinie; adres do korespondencji: ul. Bursaki 12, 20-150 Lublin.

¹ Paul Amato, przeglądając bazę danych SOCIOFILE z lat 1990-1999 zidentyfikował ponad 9 tys. publikacji, w których pojawił się termin *divorce* (por. P. A m a t o, *The Consequences of Divorce for Adults and Children*, „Journal of Marriage and the Family” 2000, nr 4). Jeżeli przyjmiemy tę liczbę za podstawę szacunków, to dochodzimy do wniosku, iż w ciągu 20 lat objętych badaniami takich publikacji pojawiło się ok. 20 tys. Analiza wszystkich tych tekstów byłaby zadaniem karkołomnym. Jakimś rozwiązaniem jest przeprowadzenie losowej selekcji artykułów. Jednak w tym przypadku losowość niekoniecznie doprowadziłaby nas do uzyskania wyboru tekstów najlepiej reprezentujących metodologiczny dorobek w interesującym

WIELOKROTNA REGRESJA LINIOWA

Wielokrotna regresja liniowa jest najczęściej stosowaną metodą analizy statystycznej w badaniach nad rozwodami². Metoda ta pozwala budować modele zależności liniowej między wieloma zmiennymi (dokładniej: jedną zmienną niezależną i wieloma zmiennymi niezależnymi). Rezultatem jest model empiryczny w postaci równania regresji. Równanie regresji ma postać: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e$, gdzie parametr e oznacza błąd. Znając wartość błędu, znalibyśmy wszystkie parametry równania, a więc moglibyśmy przewidywać ze 100% pewnością. Ponieważ wartości błędu nie znamy, przedstawiony model jest modelem teoretycznym (model empiryczny ma postać: $Y = a + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$).

Jak wynika to z samej nazwy, regresja liniowa wymaga przyjęcia założenia, że zależność między zmiennymi jest liniowa. W praktyce słuszność takiego założenia jest prawie niemożliwa do udowodnienia, jednak na szczęście procedury regresji wielorakiej są dość odporne na niewielkie odstępstwa od tego założenia. Zawsze godną polecenia regułą powinno być przeanalizowanie pod tym kątem dwuwymiarowych wykresów rozrzutu badanych zmiennych.

Chociaż większości założeń związanych z regresją wielokrotną nie można sprawdzić bezpośrednio, to poważne odstępstwa od nich dają się wykryć i powinny być odpowiednio potraktowane. W szczególności odstające obserwacje (tzn. przypadki ekstremalne) mogą poważnie zaburzać wyniki przez „przesuwanie” linii regresji w jakimś kierunku, powodując zmianę wartości współczynników regresji. Czasami przez usunięcie tylko jednego odstającego

nas zakresie. Takie założenie w znacznym stopniu zaważyło nad wyborem strategii doboru materiałów: skoncentrowano się głównie na publikacjach w renomowanych czasopismach amerykańskich, wiodących prym w zakresie socjologicznych badań nad rozwodami. Do takich należą: „Journal of Marriage and the Family”, „Family Relations”, „Journal of Divorce” (przemianowany na „Journal of Divorce & Remarriage” i „Journal of Family Issues”. Uwzględniono też teksty opublikowane w innych renomowanych czasopismach, w których okazjonalnie pojawiają się socjologiczne publikacje o rozwodach: „Sociological Spectrum”, „Social Forces”, „American Sociological Review”, „Social Indicators Research”, „American Journal of Sociology”, „Social Problems”. Ponieważ badania nad rozwodami są uprawiane również z perspektywy interdyscyplinarnej (najczęściej z pogranicza socjologii i psychologii lub socjologii i demografii), wykorzystano też publikacje ukazujące się w innych, uznanych czasopismach naukowych.

² Na podstawie szacunków można stwierdzić, iż mniej więcej połowa wszystkich analiz wielozmiennych stosowanych w badaniach nad rozwodami posługuje się wielokrotną analizą regresji. Jeżeli do tego dodać analizy wykorzystujące różne metody analizy, w tym także regresję, ten odsetek będzie jeszcze większy.

punktu można uzyskać kompletnie różne wyniki analizy. Stąd niezbędne staje się rozpoznanie i (czasowe) usunięcie przypadków ekstremalnych ze zbioru danych.

W badaniach nad rozwodami regresję wielokrotną wykorzystuje się w celu zbadania wpływu szeregu zmiennych niezależnych na zmienną zależną. Regresja pozwala na ustalenie hierarchii wpływu zmiennych, czyli pozwala stwierdzić, które zmienne niezależne wywierają najsilniejszy wpływ na zmienną zależną (w określonym zestawie zmiennych). Regresja umożliwia także oszacowanie dopasowania modelu do danych (obliczenie procentu wariancji wyjaśnionej), ewentualnie porównanie modeli konkurencyjnych. Umożliwia także predykcję na podstawie utworzonego równania regresyjnego³.

Najważniejsze wyniki regresji to (jako minimum): (a) stała regresji – wartość wyrazu wolnego w równaniu regresji (β_0) oraz współczynniki regresji cząstkowej – wartości wyrazów związanych ze zmiennymi; (b) standaryzowany współczynnik regresji (beta), wyrażony w wielkościach standaryzowanych umożliwia porównywanie wpływu poszczególnych zmiennych ujętych w równaniu; (c) współczynnik determinacji R^2 , miara dopasowania modelu liniowego do danych; współczynnik determinacji może być interpretowany jako część wariancji zmiennej zależnej wyjaśniona przez model regresji (wariancja wyjaśniona); (d) skorygowany współczynnik determinacji (R^2 skorygowany), precyzyjnie odzwierciedla dopasowanie modelu do danych. Zaletą skorygowanego R^2 jest mniejsza wrażliwość na liczbę zmiennych w równaniu. R^2 rośnie z każdą wprowadzoną nową zmienną, lecz nie oznacza to wcale, że zwiększając liczbę zmiennych w modelu, uzyskujemy coraz lepsze jego dopasowanie. Skorygowany R^2 nie zależy od liczby zmiennych i nie wzrasta wraz z wprowadzaniem nowych zmiennych do modelu, jeżeli zmienne te nie poprawiają dopasowania modelu.

Jednym z warunków wstępnych prawidłowego stosowania regresji jest numeryczny (interwałowy lub ilorazowy) poziom pomiaru wszystkich zmiennych⁴. Jest to poważne ograniczenie, gdyż wiele zmiennych w naukach spo-

³ Podstawowym ograniczeniem metodologicznym leżącym u podstaw wszystkich technik regresyjnych jest fakt, że za ich pomocą można jedynie przekonać się o istnieniu relacji, nie można natomiast dowieść istnienia związku przyczynowego będącego podłożem tej relacji.

⁴ Według dokumentacji pakietu SPSS for Windows „zmienne zależne i niezależne powinny być zmiennymi ilościowymi. Zmienne jakościowe muszą być zakodowane w zmiennych binarnych (sztucznych) lub innych zmiennych kontrastowych. Dla każdej wartości zmiennej niezależnej rozkład zmiennej zależnej musi być normalny. Wariancja rozkładu zmiennej zależnej powinna być stała dla wszystkich wartości zmiennej niezależnej. Relacja między zmienną

łecznych ma charakter nieciągły. Niekiedy jednak udaje się wprowadzić do równania regresji zmienną nominalną, stosując tzw. kodowanie zerojedynkowe⁵. Wielokrotna regresja liniowa ze skalowaniem optymalnym jest modyfikacją pozwalającą na wprowadzenie zmiennych nominalnych i porządkowych, które podlegają transformacji do postaci zmiennych numerycznych i dopiero wtedy są wprowadzane do analizy. Jest to rozwiązanie bardzo dogodne, pozwalające ominąć meandry kodowania zerojedynkowego⁶. Dziwi więc fakt, że jest to możliwość wykorzystywana bardzo rzadko. Autorowi niniejszego opracowania nie udało się znaleźć w literaturze dotyczącej rozwodów przykładu stosowania regresji ze skalowaniem optymalnym, poza swoim własnym tekstem, który dotyczył postaw wobec rozwodów⁷.

Przykładem typowego wykorzystania regresji wielokrotnej w badaniach nad rozwodami mogą być badania dotyczące zależności między zmiennymi opisującymi historię małżeństwa i zmienną akceptacji rozwodu⁸. Grupa zmiennych niezależnych obejmowała m.in. następujące zmienne: przyjaźń, ekspresję emocjonalną, harmonię małżeńską, długość trwania małżeństwa, lata edukacji, posiadanie alternatywnego partnera. Zmienną zależną była akceptacja rozwodu. Zmienna ta (poziom pomiaru interwałowy) została utworzona na podstawie odpowiedzi respondentów na 15 pytań kodowanych zerojedynkowo. Analizę przeprowadzono osobno dla kobiet i mężczyzn. Objęła ona: wskazanie zmiennych niezależnych oddziałujących na zmienną zależną, ukazanie hierarchii wpływu tych zmiennych (beta) oraz interpretację wartości skorygowanego współczynnika determinacji (jako procent wariacji wyjaśnionej). Wnioski zostały wyprowadzone prawidłowo. Zmienne wprowadzone do modelu mierzone były na poziomie numerycznym (zmienne oryginalne lub odpowiednio przekształcone). Zaskakujące jest jednak to, iż zastosowano metodę wnioskowania statystycznego w sytuacji nielosowego doboru próby.

zależną a każdą zmienną niezależną powinna być liniowa, a wszystkie obserwacje powinny być niezależne”.

⁵ Por. J. B r z e z i ń s k i, *Metodologia badań psychologicznych*, Warszawa 1997, s. 370 n.

⁶ Kodowanie zerojedynkowe wymaga dychotomizacji zmiennej, a więc redukcji poziomu pomiaru, co zawsze wiąże się z utratą większej lub mniejszej porcji informacji.

⁷ P. R y d z e w s k i, *Rozwód w opinii społecznej w latach 1990*, „Studia Demograficzne” 1999, nr 1.

⁸ L. T h o m p s o n, G. B. S p a n i e r, *The End of Marriage and Acceptance of Marital Termination*, „Journal of Marriage and the Family” 1983, Feb.

Wielokrotna analiza regresji została wykorzystana w badaniach ekonomicznych konsekwencji rozwodu. Przeprowadzono badania porównawcze osób rozwiedzionych oraz osób pozostających w związkach małżeńskich. Zmienną zależną był dochód przy uwzględnieniu wpływu szeregu zmiennych kontrolnych⁹. W innych badaniach porównawczych wyodrębniono dwie grupy rozwiedzionych ojców – mieszkających w Stanach Zjednoczonych i w Izraelu. Przedmiotem badania były relacje między byłymi małżonkami, relacje dzieci z ojcem oraz różne typy ojcowskiej aktywności w tym zakresie (wizyty, rozmowy telefoniczne, itd.). Analiza polegała m.in. na porównaniu odsetka wariancji wyjaśnionej (*nota bene* – dość wysokiego, sięgającego 27% w przypadku modelu amerykańskiego i aż 40% w przypadku modelu izraelskiego)¹⁰.

Niekiedy jednym z głównych celów analizy wykorzystującej regresję wielokrotną jest przetestowanie pewnej liczby modeli, które następnie są porównywane. W jednym z badań analizę regresji zastosowano do przetestowania dwóch modeli związków między zmiennymi – oddzielnie dla ojców i matek. Celem analizy było ukazanie powiązań między poziomem konfliktu w porozwodowych kontaktach między byłymi małżonkami i ich „definicjami sytuacji”, w szczególności – zdolności do definiowania byłego partnera jako współpartnera (w zakresie wychowania dziecka) – nie zaś jako byłego małżonka¹¹.

W badaniach nad rozwodami wielokrotna analiza regresji często wykorzystywana jest jako jedna z kilku metod analizy danych (w ramach tego samego projektu badawczego). Na przykład w badaniach nad intergeneracyjną transmisją rozwodów regresja wielokrotna była jedną z trzech metod analizy – posłużyła do porównania dwóch modeli, których celem było wyjaśnienie powiązań między rozwodem rodziców (żony, męża lub obojga) i problemami w relacjach interpersonalnych małżonków, które zdefiniowano w kategoriach zachowań zwiększających ryzyko rozpadu związku¹².

⁹ A.L. Stroup, G.E. Pollock, *Economic Well-Being Among White Elderly Divorced*, „Journal of Divorce and Remarriage” 1999, No 3/4.

¹⁰ P. I. E r e r a, C. M i n t o n, K. P a s l e y, S. M a n d e l, *Fathering After Divorce in Israel and the U.S.*, „Journal of Divorce and Remarriage” 1999, No 1/2.

¹¹ D.A. M a d d e n - D e r d i c h, S.A. L e o n a r d, F.S. C h r i s t o p h e r, *Boundary Ambiguity and Coparental Conflict after Divorce: An Empirical Test of Family Systems Model of the Divorce Process*, „Journal of Marriage and the Family” 1999, Aug.

¹² P.R. A m a t o, *Explaining the Intergenerational Transmission of Divorce*, „Journal of Marriage and the Family” 1996, Aug.

W badaniach nad strategiami przystosowania do rozwodu regresja wielokrotna została wykorzystana w celu zbadania związku między czynnikami (zidentyfikowanymi dzięki uprzednio zastosowanej analizie czynnikowej) i zmiennymi zależnymi: zaburzeniami nastroju oraz poczuciem satysfakcji z życia. Tak więc – w odróżnieniu od przykładu opisanego powyżej – regresja była jedną z metod wykorzystywanych sekwencyjnie (wyniki uzyskane dzięki zastosowaniu jednej metody analizy posłużyły jako punkt wyjścia dla kolejnej metody analizy)¹³.

Przykłady zastosowania regresji liniowej w badaniach nad rozwodami można byłoby mnożyć. Z czego wynika popularność tej metody analizy danych? Nasuwają się dwie możliwe odpowiedzi: (1) model korelacyjno-regresyjny jest stosunkowo prosty (w porównaniu z bardziej złożonymi, np. korelacją kanoniczną, modelem LISREL) zarówno na etapie konceptualizacji, jak też interpretacji wyników, (2) w badaniach socjologicznych przyjęła się pewnego rodzaju tradycja (ze względu na długotrwałość powodzenia trudno to określić mianem mody) wykorzystania wielokrotnej regresji liniowej. Tradycja ta jest podtrzymywana i utrwalana wśród kolejnych pokoleń socjologów głównie za sprawą podręczników do metod badań¹⁴. Również piszący te słowa często wykorzystuje wielokrotną regresję liniową w swoich analizach i zachęca do stosowania tej metody studentów.

REGRESJA LOGISTYCZNA I MODELE PROBIT

Regresja logistyczna, inaczej określana jako model logarytmiczno-liniowy lub w skrócie „logit”, jest analizą użyteczną w sytuacjach, w których celem jest przewidywanie obecności lub braku określonej cechy na podstawie pewnej liczby zmiennych niezależnych. Regresja logistyczna może być zastosowana w analizach, w których zmienna zależna jest dychotomiczna (nominalna). Pozwala obliczyć prawdopodobieństwo wystąpienia określonego zdarzenia. Procedurą analityczną bardzo zbliżoną do regresji logistycznej są modele

¹³ W.H. B e r m a n, D.C. T u r k, *Adaptation to Divorce: Problems and Coping Strategies*, „Journal of Marriage and the Family” 1981, Feb.

¹⁴ Np. Ch. F r a n k f o r t - N a c h m i a s, D. N a c h m i a s, *Metody badawcze w naukach społecznych*, Poznań 2001, s. 459: „Technika regresji wielokrotnej jest najbardziej właściwą techniką, którą można zastosować do analizowania problemów, gdzie występują dwie lub więcej zmiennych”.

probit. Podobnie jak w modelach logarytmiczno-liniowych zmienna zależna jest dychotomiczna. W przypadku większości pytań badawczych z zakresu socjologii rodziny i badań nad rozwodami wybór modelu logit lub probit jest jedynie kwestią dostępności odpowiedniego oprogramowania¹⁵.

Modele logit i probit są często wykorzystywane w badaniach nad rozwodami. Na przykład regresja logistyczna została wykorzystana w badaniach nad źródłami problemów małżeńskich i jego rozpadem. Małżeńskie problemy zostały potraktowane jako zbiór predyktorów dla rozpadu związku. Najbardziej interesująca część analizy polegała na testowaniu dwóch modeli wyjaśniających rozpad małżeństwa: model pierwszy – był oparty na szeregu zmiennych społeczno-demograficznych, o których wiadomo (na podstawie wcześniejszych badań), że są powiązane z prawdopodobieństwem rozwodu (wiek zawarcia małżeństwa, rozwód w rodzinie pochodzenia, długość trwania małżeństwa, poziom wykształcenia, religijność itd.). Model drugi zawierał zmienne związane z relacjami małżeńskimi i stylem życia małżonków (zazdrość, zmienność nastrojów, niewierność, irytujące nawyki, spożycie alkoholu i używanie narkotyków itp.). W obrębie każdego z tych modeli poszukiwano zmiennych będących najsilniejszymi predyktorami rozwodu¹⁶.

W badaniach nad wpływem socjoekonomicznego statusu społeczności lokalnej na proces rozpadu małżeństw wzięto pod uwagę wiele zmiennych opisujących sąsiedztwo osób rozwiedzionych: procent mieszkańców posiadających wysokie/niskie dochody, wysoki/niski poziom wykształcenia, wysoki/niski status zawodowy itp. Wśród zmiennych znalazły się między innymi zmienne opisujące demograficzny i socjoekonomiczny status żon (aktywność zawodowa, poziom wykształcenia oraz wiek żony w momencie zawierania małżeństwa) i mężów (poziom wykształcenia, status zawodowy oraz wysokość dochodów w skali roku), a także cechy małżeństwa jako całości (trwałość/rozpad, staż małżeński, posiadanie własnego domu itp.). Zaprezentowano trzy modele regresji logistycznej: model pierwszy – pełny zbiór zmiennych z wyjątkiem złożonej zmiennej NDI (*Neighborhood Disadvantage Index*), opartej na wskaźniku ubóstwa, odsetku osób korzystających z pomocy społecznej, poziomu bezrobocia itp. Model drugi zawierał wyłącznie zmienną NDI, model trzeci – wszystkie zmienne. Regresja logistyczna pozwoliła osza-

¹⁵ S.P. Morgan, J.D. Teachman, *Logistic Regression: Descriptions, Examples, and Comparisons*, „Journal of Marriage and the Family” 1988, Nov.

¹⁶ P.R. Amato, S.J. Rogers, *A Longitudinal Study of Marital Problems and Subsequent Divorce*, „Journal of Marriage and the Family” 1997, Aug.

cować każdy z powyższych modeli i rozstrzygnąć postawione hipotezy badawcze¹⁷.

Modele logit i probit najczęściej są wykorzystywane w tych badaniach nad rozwodami, w których dominuje problematyka demograficzna i ekonomiczna. Wynika to prawdopodobnie z możliwości uwzględnienia dużej liczby zmiennych zawartych w ogólnokrajowych próbach badawczych, wykorzystywanych najczęściej w ramach przyjętej problematyki oraz strategii analitycznej. Bardzo szeroki zestaw zmiennych wykorzystali badacze stosujący model logarytmiczno-liniowy w badaniach nad wpływem na stabilność małżeństwa relatywnych wielkości dochodów uzyskiwanych z pracy przez mężów i żony (łącznie ponad 30 zmiennych niezależnych). Szczególnie interesująca okazała się ekonomiczna analiza czterech modeli relacji dochodów: tradycyjnego (zarobki męża stanowią 75-100% dochodów rodziny), nowego – tradycyjnego (odpowiednio: 50-75%), nietradycyjnego (udział zarobków żony w dochodach rodziny w granicach 50-75%) i odwróconego – tradycyjnego (żona dostarcza 75-100% dochodów rodziny)¹⁸. Podobną problematykę podjął inny badacz, również wykorzystujący regresję logistyczną jako podstawową metodę analizy danych. Zbiór zmiennych niezależnych został znacznie rozszerzony – do ponad 40. Utworzono siedem modeli zawierających różne zestawy zmiennych. Pozwoliło to na sformułowanie bardzo interesujących wniosków, stawiających analizowany problem w całkowicie nowym świetle¹⁹.

Przykładem zastosowania modelu probit jest badanie powiązań między ryzykiem rozpadu związku i wczesnym wiekiem zawarcia małżeństwa oraz wczesnym rodzicielstwem. Przetestowano następujące modele: (1) wczesny wiek zawarcia małżeństwa, (2) wczesne rodzicielstwo oraz (3) wczesny wiek zawarcia małżeństwa i wczesne rodzicielstwo. Analizy przeprowadzono oddzielnie dla białych i czarnych²⁰.

W innych badaniach wykorzystano model probit w celu oszacowania prawdopodobieństwa separacji lub rozwodu w małżeństwach białych kobiet (w cią-

¹⁷ S.J. S o u t h, *The Geographic Context of Divorce: Do Neighborhoods Matter?*, „Journal of Marriage and the Family” 2001, Aug.

¹⁸ D.A. H e c k e r t, T.C. N o w a k, K.A. S n y d e r, *The Impact of Husbands' and Wives' Relative Earnings on Marital Disruption*, „Journal of Marriage and the Family” 1998, Aug.

¹⁹ H. O n o, *Husbands' and Wives' Resources and Marital Dissolution*, „Journal of Marriage and the Family” 1998, Aug.

²⁰ K.A. M o o r e, L.J. W a i t e, *Marital Dissolution, Early Motherhood and Early Marriage*, „Social Forces” 1981, Sept.

gu pierwszych pięciu lat trwania związku oraz w czasie kolejnej „pięcioletki”), wiążąc ryzyko rozpadu małżeństwa z seksualnymi doświadczeniami kobiet w okresie przedmałżeńskim. Przetestowano sześć modeli z różną kombinacją zmiennych²¹.

Modele logit i probit oferują wiele możliwości w zakresie badań nad rozwodami, pozwalają na wprowadzenie zmiennej dychotomicznej jako zmiennej zależnej. W badaniach nad rozwodami jest to najczęściej zmienna (różnie nazywana) o wartościach: małżeństwo – rozwód. Celem jest więc określenie zmiennych, które powodują zwiększenie ryzyka rozwodu. W praktyce sprowadza się to do testowania różnorodnych modeli, różniących się zestawem zmiennych niezależnych. Jak już wspomniano wyżej, badania wykorzystujące modele logit i probit to najczęściej analizy wtórne, korzystające z gotowych zbiorów danych, najczęściej powstających wskutek realizacji dużych, ogólnokrajowych projektów badawczych. Niekiedy można odnieść wrażenie, iż badacze nie są w stanie oprzeć się pokusie wprowadzenia do analizy wszystkich dostępnych zmiennych, które wiążą się wprost lub pośrednio z badanym zagadnieniem. Stąd w analizach danych opartych na regresji logistycznej i modelach probit bardzo często spotykamy się z bardzo dużymi zestawami, liczącymi po 30-40 zmiennych niezależnych. Niekiedy sprawia to wrażenie swoistej „żonglerki” zmiennymi.

ANALIZA KORELACYJNA

W badaniach nad rozwodami analiza korelacji rzadko wykorzystywana jest jako główna lub jedyna metoda analizy danych, natomiast stosunkowo często jako jedna z metod.

Korelacja jest miarą powiązania pomiędzy dwiema lub większą liczbą zmiennych. Przed przystąpieniem do obliczania współczynnika korelacji należy przejrzeć dane w poszukiwaniu wartości skrajnych (które mogą powodować błędne wyniki) i dowodów zależności liniowej.

W praktyce najczęściej stosowany jest współczynnik korelacji Pearsona, który jest miarą skojarzeń liniowych. Dwie zmienne mogą być idealnie powiązane, lecz jeśli relacja nie jest liniowa, współczynnik korelacji Pearsona

²¹ J.R. Kahn, K.A. London, *Premarital Sex and the Risk of Divorce*, „Journal of Marriage and the Family” 1991, Nov.

nie jest odpowiednią statystyką. Dla danych jakościowych można stosować współczynnik Phi, V Cramera, współczynnik kontyngencji, lambda oraz tau Goodmana i Kruskala. Gdy zmienne mierzone są na skali porządkowej, można stosować współczynnik gamma, tau-b Kendalla i tau-c Kendalla²².

Przykładem badań, w których podstawową (jedyną) metodą analizy danych była korelacja, jest empiryczne testowanie Lewisa i Spaniera teorii jakości i stabilności małżeństwa, rozszerzonej o zmienne przystosowania porozwodowego. Wykorzystując współczynnik korelacji r Pearsona, zbadano związek między poziomem przystosowania porozwodowego i (a) wskaźnikami jakości małżeństwa, (b) wskaźnikami alternatyw wobec (byłego) małżeństwa i (c) siłą barier wobec rozwodu. Współczynniki korelacji zostały policzone dla całej próby oraz oddzielnie dla kobiet i mężczyzn. Wszystkie zmienne były numeryczne²³.

Przykładem innych badań, w których korelacja stanowiła główne narzędzie analityczne, było studium powiązań między cechami społeczno-demograficznymi osób rozwiedzionych, definiowanymi przez nie obszarami niepowodzeń w (byłym) małżeństwie i przejawami zaburzeń psychicznych²⁴. W analizach wykorzystano współczynnik korelacji *tau-b* Kendalla²⁵.

Interesującym przykładem jest badanie problemu naznaczania (stygmatyzacji) osób rozwiedzionych. Podjęto próbę sprawdzenia, jak silny jest związek między odsetkiem przyjaciół osób rozwiedzionych, które wyraziły dezaprobatę wobec rozwodu i zespołem cech społeczno-demograficznych rozwiedzionych²⁶. Dodatkowo uwzględniono także fakt ewentualnego romansu, który miał miejsce w trakcie trwania małżeństwa. Ten interesujący projekt badawczy ma jednak przynajmniej dwie istotne wady. Po pierwsze – próba nie zo-

²² Szczegółowe zestawienie najważniejszych współczynników korelacji według poziomu pomiaru zmiennych można znaleźć np. w: R. M a y n t z, K. H o l m, P. H u e b n e r, *Wprowadzenie do metod socjologii empirycznej*, Warszawa 1985, s. 250.

²³ R.G. G r e e n, *The Influence of Divorce Prediction Variables on Divorce Adjustment: An Expansion and Test of Lewis' and Spanier Theory of Marital Quality and Marital Stability*, „Journal of Divorce” 1983, nr 1.

²⁴ G.C. K i t s o n, M.B. S u s s m a n, *Marital Complaints, Demographic Characteristics, and Symptoms of Mental Distress in Divorce*, „Journal of Marriage and the Family” 1982, Feb.

²⁵ *Tau* Kendalla to nieparametryczna miara zależności dla zmiennych porządkowych lub rangowanych, uwzględniająca powiązania rang. Znak współczynnika wskazuje na kierunek zależności, a jego wartość bezwzględna ukazuje siłę związku. Większe wartości bezwzględne wskazują na silniejsze zależności. Współczynnik przyjmuje wartości z zakresu od -1 do +1. Jednak wartości -1 lub +1 mogą zostać otrzymane jedynie dla tabel kwadratowych.

²⁶ N. G e r s t e l, *Divorce and Stigma*, „Social Problems” 1987, nr 2.

stała dobrana metodą losową; blisko 40% respondentów zostało dobranych w sposób celowy: częściowo akcydentalny, częściowo przy wykorzystaniu strategii „kuli śniegowej”. A przecież analiza korelacji jako metoda wnioskowania statystycznego wymaga doboru losowego próby. Po drugie, nie podano, jaki współczynnik korelacji został zastosowany. Sytuację komplikuje fakt, iż uwzględnione zmienne są mierzone na różnych poziomach pomiaru (nominalne, porządkowe i ilorazowe), a więc można byłoby oczekiwać wykorzystania różnych współczynników korelacji.

W innych badaniach nad małżeństwami, które zdecydowały się na rozwód, lecz ostatecznie wycofały pozew, wykorzystano korelację cząstkową, która umożliwia wyliczenie współczynnika opisującego liniową relację między dwiema zmiennymi, w zależności od jednej lub wielu dodatkowych zmiennych. Jest to korelacja, która pozostaje między dwiema zmiennymi po usunięciu korelacji będącej skutkiem istnienia współzależności z innymi zmiennymi. Tak więc korelacja cząstkowa to korelacja pomiędzy zmienną zależną i zmienną niezależną, gdy efekty liniowe pozostałych zmiennych niezależnych w modelu zostały usunięte dla obu zmiennych²⁷.

W omawianych badaniach uwzględniono obszerny zbiór zmiennych (41), w tym między innymi: konsumpcję alkoholu, przemoc wobec współmałżonka, niektóre cechy psychologiczne, relacje małżeńskie, sytuację zdrowotną, cechy społeczno-demograficzne. Powyższe zmienne były korelowane ze statusem małżeńskim (grupa małżonków, którzy wycofali pozwy i grupa rozwiedzionych lub pozostających w separacji), który przekodowano zerojedyńkowo (umożliwiło to zastosowanie korelacji r Pearsona). Przeprowadzono dwie równoległe analizy korelacji: r Pearsona (stopnia zerowego) i korelację cząstkową (stopnia trzeciego), kontrolując wpływ płci, wieku i rasy²⁸.

Analiza korelacyjna często jest wykorzystywana wraz z analizą regresji²⁹. W badaniach nad rozwodami można znaleźć liczne przykłady takiego podejścia. Na przykład w badaniach nad stopniem i uwarunkowaniami zaangażowania ojców w wypełnianie ról rodzicielskich w pierwszym roku po rozwodzie

²⁷ W odróżnieniu od korelacji cząstkowej, która w zależności od liczby dodatkowych zmiennych może być korelacją stopnia pierwszego, drugiego itd. – „zwykłą” korelację niekiedy określa się jako korelację stopnia zerowego (zero – order).

²⁸ G.C. K i t s o n, J.K. L a n g l i e, *Couples Who File for Divorce but Change Their Minds*, „American Journal of Orthopsychiatry” 1984, nr 3.

²⁹ Np. J. Brzeziński używa terminu *model korelacyjno-regresyjny*; por. J. B r z e z i ń - s k i (red.), *Wielozmiennowe modele statystyczne w badaniach psychologicznych*, Warszawa–Poznań 1987, s. 12.

skorelowano: poziom konfliktu między byłymi małżonkami, zakres współpracy w zakresie wykonywania obowiązków rodzicielskich, relacje dziecko–rodzic, poziom satysfakcji z pełnienia ról rodzicielskich oraz intensywność interakcji między byłymi małżonkami w zakresie pełnienia ról rodzicielskich. Analiza korelacyjna została następnie uzupełniona hierarchiczną regresją liniową³⁰.

W badaniach poświęconych związkom między rozpadem małżeństw i wczesnym wiekiem ich zawierania skorelowano: status społeczno-ekonomiczny rodziny pochodzenia, grupę zmiennych opisujących kulturowe korzenie małżonków, grupę zmiennych charakteryzujących strukturę rodziny pochodzenia oraz przynależność do określonej kohorty urodzenia – z poziomem wykształcenia i stabilnością małżeństwa (rozbite–trwałe)³¹. Po analizie korelacyjnej przeprowadzono analizę regresji. Niestety, w opisie procedury nie znajdujemy informacji, który ze współczynników korelacji został obliczony.

ANALIZA WARIANCJI

W badaniach nad rozwodami analiza wariancji jest wykorzystywana znacznie rzadziej niż regresja liniowa. Celem analizy wariancji jest testowanie różnic między średnimi. Jeżeli porównywane są jedynie dwie średnie, analiza wariancji daje takie same rezultaty jak test t. Tak więc, analiza wariancji jest wykorzystywana wtedy, gdy: (a) porównujemy trzy grupy lub więcej, (b) zmienne niezależne (główne) są jakościowe, (c) zmienna zależna jest numeryczna.

Analiza wariancji może mieć postać dwuzmienną (jednak zmienna niezależna i jednak zmienna zależna) lub wielozmienną (więcej niż jedna zmienna niezależna i jedna zmienna zależna). W odróżnieniu od wersji wielozmiennej (ANOVA) wersja dwuzmienna (jednoczynnikowa) określana się często jako ONE-WAY. To rozróżnienie terminologiczne jednak nie jest rygorystycznie przestrzegane, dlatego też można natrafić na liczne przykłady

³⁰ D.A. M a d d e n - D e r d i c h, S.A. L e o n a r d, *Parental Role Identity and Fathers' Involvement in Coparental Interaction After Divorce: Fathers' Perspectives*, „Family Relations” 2000, nr 3.

³¹ D.D. W i t t, B. D a v i d s o n, D.L. S o l l i e, G.D. L o w e, C.W. P e e k, *The Consequences of Early Marriage on Marital Dissolution*, „Sociological Spectrum” 1987, nr 7.

używania terminu *analiza wariancji* lub *ANOVA* również w przypadku analiz dwuzmiennowych.

Jednoczynnikową analizę wariancji wykorzystuje się do testowania hipotezy o równości kilku średnich. Technika ta jest rozszerzeniem testu *t* dla jednej próby. Oprócz wykazania różnic między średnimi może zaistnieć potrzeba określenia, które średnie są różne. Istnieją dwa typy testów służących do porównywania średnich: kontrasty *a priori* i testy *post hoc*. Kontrasty są testami zdefiniowanymi przed przeprowadzeniem badania (najczęściej eksperymentu), natomiast testy *post hoc* są przeprowadzane po zakończeniu badań.

W jednoczynnikowej analizie wariancji (ONE-WAY) obserwowana zmienność w próbie składa się z dwóch komponentów: wariancji wewnątrzgrupowej i wariancji międzygrupowej. Wariancja wewnątrzgrupowa jest sumą kwadratów odchyłeń poszczególnych obserwacji od średniej z każdej grupy. Wariancja międzygrupowa jest sumą kwadratów odchyłeń średnich z grup od średniej ogólnej. Ponieważ suma kwadratów zależy od liczby obserwacji (tzn. przypadków w grupach oraz grup w próbie), oblicza się średnią sumę kwadratów. Wynik testu *F* jest stosunkiem międzygrupowej średniej sumy kwadratów do wewnątrzgrupowej średniej sumy kwadratów. *F* jest tym większe, im większa jest wariancja między grupami oraz im mniejsza jest wariancja wewnątrz grup.

Jednoczynnikowa analiza wariancji została wykorzystana na przykład w badaniach, których celem było porównanie poziomu depresji u kobiet rozwiedzionych, wdów oraz zamężnych. Badania zrealizowano w Izraelu na grupie kobiet matek³². Analizę tę uzupełniono wielokrotną regresją liniową.

W badaniach nad zdrowotnymi skutkami wczesnych faz procesu rozwodowego wykorzystano jednoczynnikową analizę wariancji w celu porównania: (a) małżonków pozostających w separacji i małżonków mieszkających razem, (b) małżonków, którzy rozważali możliwość rozwodu i małżonków, którzy takiego rozwiązania nie rozważali³³. Oprócz analizy wariancji zastosowano również test *t* oraz test chi-kwadrat.

Gdy już zostanie wykazane istnienie różnic między średnimi, za pomocą testów wielokrotnych porównań można określić, które średnie są różne. Wie-

³² R. K a t z, *Marital Status and Well-Being: A Comparison of Widowed, „Divorced and Married Mothers in Israel, Journal of Divorce”* 1991, nr 3/4.

³³ G.C. K i t s o n, A.V. G r a h a m, D.D. S c h m i d t, *Troubled Marriages and Divorce: A Prospective Suburban Study, „Journal of Family Practice”* 1983, nr 2.

lokrotne porównania testują różnicę między każdą parą średnich i generują macierz, na której zaznaczone są znacząco różne średnie grupowe. Zastanawiające jest, iż w żadnym znanym przypadku badań nad rozwodami, w których zastosowano jednoczynnikową analizę wariancji nie została ona uzupełniona testem porównań wielokrotnych. Jest to poważny mankament, gdyż ogólny wynik testu F nie daje wglądu w rzeczywistą strukturę różnic między średnimi. W skrajnym przypadku może dojść do sytuacji, w której bardzo duża różnica między dwiema średnimi spowoduje, że test F okaże się istotny przy braku różnic między pozostałymi średnimi.

Celem analizy wariancji wielozmiennowej (ANOVA) jest wykrycie wkładu poszczególnych czynników do zaobserwowanej całkowitej wariancji (zmienności) zmiennej zależnej (tzw. dekompozycja sumy kwadratów). Na wariancję całkowitą składa się wariancja wywołana przez zmienne niezależne (każdą oddzielnie), wariancja wywołana przez interakcję zmiennych niezależnych, wariancja powodowana przez błąd (wariancja wewnątrzgrupowa, której przyczyny nie są znane).

Po ustaleniu wkładu poszczególnych czynników do wariancji całkowitej należy zbadać, czy wpływ zmiennych niezależnych na zmienną zależną jest istotny, tzn. czy odzwierciedla istniejący faktycznie trend czy też jest skutkiem przypadku (obliczenie wartości statystyki F i określenie poziomu istotności), a także – jak silny jest wpływ zmiennych niezależnych na zmienną zależną (obliczenie współczynnika eta kwadrat i współczynnika omega).

Model ANOVA może być zaakceptowany wtedy, gdy wariancja wyjaśniona jest istotna. Poziom istotności wariancji wyjaśnionej jest odpowiednikiem współczynnika determinacji R^2 w analizie regresji.

W badaniach nad rozwodami ANOVA została wykorzystana w analizach, których celem było określenie poziomu przystosowania porozwodowego w różnych sferach życia społecznego (rodzicielstwo, zatrudnienie, kwestie finansowe, relacje społeczne itp.). Zmiennymi niezależnymi były płeć oraz czas, który upłynął od momentu rozwodu (jako zmienna dychotomiczna: bezpośrednio po rozwodzie oraz po 2-4 latach po rozwodzie)³⁴. W innych badaniach zastosowano analizę wariancji, aby zidentyfikować czynniki, które wiążą się z uczuciami żywionymi wobec byłego współmałżonka (skala przyjacielskość-wrogość). Zmiennymi niezależnymi były: posiadanie dzieci, fakt zawarcia

³⁴ R. Katz, N. Pech, *Adjustment to Divorce in Israel: A Comparison Between Divorced Men and Women*, „Journal of Marriage and the Family” 1985, Aug.

powtórnego związku małżeńskiego i częstość kontaktów z byłym współmałżonkiem³⁵.

Interesującym przykładem wykorzystania analizy wariancji jest badanie porównawcze trzech grup osób pochodzących z rodzin: trwałych (z podziałem na trwałe-szczęśliwe i trwałe-nieszczęśliwe), rozbitych przez rozwód lub separację oraz niepełnych wskutek śmierci rodzica. Grupy te porównywano pod względem postaw wobec małżeństwa (cztery wskaźniki), wobec dzieci (dwa wskaźniki), wobec miłości romantycznej oraz wobec rozwodów (po jednym wskaźniku)³⁶. Można jednak mieć wątpliwości, czy dobór próby miał charakter probabilistyczny: w tekście nie znajdujemy odpowiedniej deklaracji na ten temat, zaś przytoczony opis procedury doboru pozwala snuć przypuszczenia, że w gruncie rzeczy dobór był nielosowy.

Podobne zastrzeżenie można postawić w stosunku do innych analiz, w których wykorzystano bardziej wyrafinowaną wersję analizy wariancji – MANOVA (więcej niż jedna zmienna zależna). W badaniach tych analizowano wpływ zbioru zmiennych społeczno-demograficznych na zbiór zmiennych reprezentujących opinie osób rozwiedzionych na temat przyczyny rozpadu ich małżeństw³⁷. W omawianych badaniach próbę badawczą stanowiła grupa rozwiedzionych wolontariuszy. Nie ma wątpliwości, iż taki sposób doboru próby nie upoważniał do stosowania metod wnioskowania statystycznego.

Analiza wariancji może być stosowana wówczas, gdy zostaną spełnione następujące warunki: (a) próby są pobrane z populacji w sposób losowy i niezależny, (b) dane pochodzą z populacji o rozkładzie normalnym, (c) wariancja we wszystkich populacjach, z których brane są próby, jest taka sama (homogeniczność wariancji).

Rozkład możemy obserwować na wykresach (plotach) danych. Matematycznie normalność rozkładu możemy zbadać np. za pomocą testu normalności Lillefora, będącego modyfikacją testu Kołmogorowa–Smirnowa. Trzeba jednak pamiętać, że dla dużych prób bardzo trudno wykazać normalność rozkładu. Jest rzeczą niemal niemożliwą znaleźć dane, które miałyby do-

³⁵ C. M a s h e t e r, *Postdivorce Relationships between Ex-spouses: the Role of Attachment and Interpersonal Conflict*, „Journal of Marriage and the Family” 1991, Feb.

³⁶ E.F. G r e e n b e r g, W.R. N a y, *The Intergenerational Transmission of Marital Instability Reconsidered*, „Journal of Marriage and the Family” 1982, May.

³⁷ A. B u r n s, *Perceived Causes of Marriage Breakdown and Conditions of Life*, „Journal of Marriage and the Family” 1984, August.

kładnie rozkład normalny. Dlatego zaakceptować można rozkład, który nie odbiega znacząco od rozkładu normalnego.

Homogeniczność wariancji bada wiele testów statystycznych. Dla analizy wariancji szczególnie użyteczny jest test homogeniczności wariancji Levene'a, mniej zależny od założenia o normalności rozkładu. Jeżeli nawet wariancje są różne, ale grupy są podobnej wielkości, nie ma powodu do niepokoju. Jeżeli jednak grupy znacznie różnią się pod względem liczebności, a wariancja nie jest homogeniczna, należy transformować dane lub stosować testy nieparametryczne.

Analizując opublikowane wyniki badań, w których wykorzystano analizę wariancji, nie znaleziono ani jednego przypadku sprawdzania powyższych założeń. Nawet jeżeli przyjąć najbardziej liberalne stanowisko w stosunku do warunków, w których ANOVA może być stosowana, oczekiwać można przynajmniej krótkiego komentarza na ten temat pokazującego, iż kwestia założeń była rozważana.

ANALIZA PRZEŻYCIA

W badaniach nad rozwodami analiza przeżycia nie jest wykorzystywana tak często jak regresja liniowa lub regresja logistyczna, niemniej jednak z pewnością należy do kanonu metod statystycznych stosowanych w ramach omawianej problematyki badawczej.

Analiza przeżycia stopniowo torowała sobie drogę wśród badań nad rozwodami. Początki stosowania tej metody analizy danych sięgają wczesnych lat osiemdziesiątych minionego wieku. Były to jednak raczej odosobnione przypadki. Dwa czynniki przyczyniły się do szerszego wykorzystania analizy przeżycia: postępująca komputeryzacja (w tym także większa dostępność odpowiedniego oprogramowania) i tendencja do poszukiwania nowych metod analizy danych w sytuacji, gdy stopniowo wyczerpywały się pomysły badawcze związane z wykorzystaniem metod bardziej tradycyjnych.

Najogólniej rzecz biorąc, należy stwierdzić, że analiza przeżycia stosowana jest w tych przypadkach, w których chcemy zbadać rozkład czasu pomiędzy dwoma wydarzeniami (np. zawarciem małżeństwa i rozwodem). Jednakże dane, którymi dysponujemy, często zawierają przypadki, dla których to drugie wydarzenie (w czasie objętym badaniem) nie zaszło. Podstawowa idea analityczna w tej sytuacji polega na podziale okresu obserwacji na krótsze interwały, w ramach których obliczane jest prawdopodobieństwo zajścia zdarzenia.

Prawdopodobieństwa oszacowane dla poszczególnych interwałów są następnie wykorzystywane do oszacowania prawdopodobieństwa wystąpienia zdarzenia w innym czasie. Analiza przeżycia (ang. *survival analysis*) zawiera następujące procedury statystyczne: tabele trwania życia, Kaplana–Meiera analizy przeżycia, regresję Coxa, regresję Coxa ze współzmienną zależną od czasu.

Regresję Coxa wykorzystano w badaniach norweskich, których celem było przetestowanie hipotezy, iż psychologiczna kondycja małżonków może służyć jako predyktor rozpadu małżeństwa. Zła kondycja psychologiczna małżonków może wynikać z problemów małżeńskich lub może być jedną z przyczyn owe problemy wywołujących. W pierwszym przypadku kondycja psychologiczna małżonków może być predyktorem rozpadu jedynie krótkoterminowo, w przypadku drugim – długoterminowo. Strategia analityczna polegała na zdefiniowaniu 8 interwałów czasu o długości 5 miesięcy każdy i zestawieniu liczby separacji w obrębie tychże interwałów. Następnie do analizy włączono zmienne kontrolne (wiek, staż małżeński, poziom urbanizacji otoczenia, liczbę dzieci, wykształcenie, dochody itp.) – w odniesieniu do wskaźnika kondycji psychologicznej i statusu małżeństwa (stabilność–rozpad) – osobno dla kobiet i mężczyzn. Kolejnym krokiem była regresja Coxa uwzględniająca wszystkie wymienione wcześniej zmienne. Analizy doprowadziły do sformułowania interesujących wniosków³⁸.

Tabele trwania życia oraz regresję Coxa (funkcję hazardu) wykorzystano w bardzo interesujących badaniach nad związkiem między poliginią i rozwodami w Nigerii. Ogół badanych małżeństw podzielono na trzy kategorie według liczby żon: monogamiczne, polygynia – dwie żony oraz poligynia – trzy żony lub więcej. Analiza przebiegała w kilku fazach: (a) obliczenie współczynników dla liczby rozwodów i liczby żon, (b) powtórna estymacja przy uwzględnieniu bezdzietności, (c) estymacja trzeciego modelu, uwzględniającego dodatkowo wpływ religijności, wykształcenia, wielkości miejscowości zamieszkania, wieku zawarcia małżeństwa i regionu kraju, (d) estymacja modelu finalnego, w którym dodatkowo uwzględniono długość trwania małżeństwa – jako zmienną, której wartość ma zostać oszacowana³⁹.

Tabele trwania życia wykorzystano również w analizie demograficznych determinantów tzw. rozwodów odłożonych (ang. *deleyed divorce*). Podstawo-

³⁸ A. M a s t e k a a s a, *Psychological Well-Being and Marital Dissolution. Selection Effects?*, „Journal of Family Issues” 1994, nr 2.

³⁹ A.J. G a g e - B r a n d o n, *The Polygyny-Divorce Relationship: A Case Study of Nigeria*, „Journal of Marriage and the Family” 1992, May.

wą grupą badanych były osoby, których staż małżeński był nie krótszy niż 10 lat. Grupą referencyjną stanowiły osoby, których staż małżeński wynosił 20 lat. Po transformacji zmiennych za punkt wyjścia przyjęto współczynniki równe 1, które zawsze charakteryzowały grupę referencyjną. Wartości odpowiednich współczynników dla pozostałych dwóch grup (15 i 25 lat pożycia) odnoszono do tej wartości. W analizach uwzględniono następujące zmienne niezależne: rasę, region, trwałość rodziny pochodzenia, wiek żony w momencie zawarcia małżeństwa, wyznanie i poziom religijności, wykształcenie męża itp.⁴⁰

Regresję Coxa (funkcję hazardu) zastosowano również w badaniach nad poziomem ryzyka rozwodu w zależności od geograficznie zdefiniowanych rynków doboru małżeńskiego, na których oboje partnerzy nawiązują kontakty społeczne, mogące przeistoczyć się w związki alternatywne wobec aktualnego małżeństwa. Wykorzystano dane z badań panelowych, budując 6 modeli i wprowadzając kolejno różne zestawy zmiennych, o których wiadomo, iż zwiększają ryzyko rozpadu małżeństwa. Znalezione powiązania między ryzykiem rozpadu i funkcjonowaniem żon w środowisku zawodowym, w którym dominują osoby płci przeciwnej. Nie potwierdzono takiej zależności w przypadku mężczyzn⁴¹.

INNE METODY ANALIZY WIELOZMIENNOWEJ

W badaniach nad rozwodami można znaleźć przykłady stosowania również mniej popularnych metod analizy: korelacji kanonicznej, analizy ścieżek, analizy czynnikowej, modelu LISREL i analizy skupień.

Korelacja kanoniczna jest rozwinięciem ogólnego modelu liniowego, którego przykładem jest np. regresja. Podobnie jak w regresji wymagany jest co najmniej interwałowy poziom pomiaru wszystkich zmiennych. Korelacja kanoniczna pozwala na badanie związku między zbiorem zmiennych niezależnych i zbiorem zmiennych zależnych. Korelacja kanoniczna nie ukazuje więc korelacji pomiędzy zmiennymi, lecz pomiędzy zbiorami zmiennych. Pozwala

⁴⁰ L.Y.W. Chan, T.B. Heaton, *Demographic Determinants of Delayed Divorce*, „Journal of Divorce” 1989, nr 1.

⁴¹ S.J. South, K. Trent, Y. Shen, *Changing Partners: Toward a Macrostructural-Opportunity Theory of Marital Dissolution*, „Journal of Marriage and the Family” 2001, Aug.

ukazać strukturę związku poprzez selekcję funkcji liniowych, które w maksymalny sposób odtwarzają korelację między zbiorem zmiennych⁴².

Korelację kanoniczną oraz wielozmienną analizę wariancji (MANOVA) wykorzystano w badaniach dotyczących związku między jakością małżeńskich interakcji (w opinii samych małżonków oraz w opinii obserwatorów zewnętrznych) i stabilnością małżeństwa⁴³. W innych badaniach celem zastosowania korelacji kanonicznej było oszacowanie ryzyka rozwodu w dwóch krytycznych (pod względem częstości rozwodów) fazach rozwoju rodziny: (1) w ciągu pierwszych 7 lat trwania związku, (2) w czasie, gdy najstarsze dziecko osiąga wiek 14 lat. Warto wspomnieć, iż (poza wstępnym wywiadem kwestionariuszowym) zastosowano rzadko wykorzystywaną w socjologii metodę eksperymentu laboratoryjnego z późniejszą analizą treści zgromadzonego materiału⁴⁴.

Analiza ścieżek jest metodą umożliwiającą weryfikację hipotez dotyczących struktury zależności przyczynowych w określonym zbiorze zmiennych (mierzonych na skali co najmniej interwałowej). Wyniki analizy ścieżek najczęściej prezentowane są w postaci tzw. diagramu ścieżek – rysunku przypominającego socjogramy stosowane w socjometrii. Niektórzy statystycy podkreślają walory analizy ścieżek, zwłaszcza w zestawieniu z tradycyjnym modelem korelacyjno-regresyjnym⁴⁵.

Analizę ścieżek wykorzystano w badaniu uwarunkowań porozwodowych więzi między byłymi małżonkami związanych z wypełnianiem ról rodzicielskich. Model zależności między zmiennymi zakładał powiązania przyczynowo-skutkowe między dwiema grupami zmiennych (należących do sfery emocjonalnej i sfery strukturalnej) i poziomem konfliktu z jednej strony oraz poziomem kooperacji – z drugiej. Poziomą więzi porozwodowej był więc wyznaczany przez konflikt i kooperację, a te czynniki przez szereg innych zmiennych. Ten teoretyczny model został przetestowany empirycznie z analizą

⁴² S.D. M c L a u g h l i n, L.B. O t t o, *Canonical Corelation Analysis in Family Research*, „Journal of Marriage and the Family” 1981, Feb.

⁴³ L.S. M t t h e w s, K.A. W i c k r a m a, R.D. C o n g e r, *Predicting Marital Instability Form Spouse and Obserwer Reports of Marital Interactions*, „Journal of Marriage and the Family” 1996, Aug.

⁴⁴ J.M. G o t t m a n, R.W. L e v e n s o n, *The Timing of Divorce: Predicting When a Couple Will Divorce Over a 14-Year Period*, „Journal of Marriage and the Family” 2000, nr 3.

⁴⁵ M. G a u l, A. M a c h o w s k i, *Elementy analizy ścieżek*, w: B r z e z i ń s k i (red.) *Wielozmiennowe modele*.

ścieżek jako główną metodą analizy danych⁴⁶. W innych badaniach analizę ścieżek zastosowano w celu określenia poziomu emocjonalnego i psychologicznego przystosowania porozwodowego. Model analizy ścieżek zakładał następujące relacje przyczynowo-skutkowe: „przedzrodzowe zasoby osobowe” każdego z małżonków (wykształcenie, zawód, niektóre cechy psychologiczne itd.) wpływają na poziom przystosowania porozwodowego – bezpośrednio oraz pośrednio – poprzez: (1) „zasoby społeczne” (sieć stosunków społecznych, posiadanie stałego partnera lub powtórne małżeństwo), (2) poziom stresu (depresja, problemy finansowe, mieszkaniowe itp.), (3) stosunek emocjonalny do byłego współmałżonka. Ogółem w modelu wyodrębniono 10 relacji przyczynowo-skutkowych. Analiza ścieżek pozwoliła na uściślenie i modyfikację modelu teoretycznego, który przedstawiono następnie w dwóch wersjach – oddzielnie dla kobiet i mężczyzn⁴⁷.

Jednakże również analiza ścieżek spotkała się z krytyką podnoszącą braki modelu. Krytyka ta sprowadza się głównie do dwóch argumentów – po pierwsze, założenia modelu wydają się nieralistyczne w odniesieniu do rzeczywistości społecznej, po drugie – wiele zmiennych pojawiających się w analizie ścieżek to nieobserwowalne, złożone struktury, które trudno powiązać (w sposób rzetelny i trafny) z pojedynczymi wskaźnikami. Problemy te skierowały uwagę badaczy na inne podejście – tzw. LISREL (ang. *Linear Structural Relationship*). LISREL łączy w sobie pewne cechy analizy czynnikowej i wielokrotnej regresji liniowej. Pozwala na estymację związków przyczynowych między nieobserwowalnymi (latentnymi) zmiennymi. Wadą metody jest jej złożoność⁴⁸.

Przykładem badań, w których zastosowano LISREL, jest analiza związków między trudną sytuacją ekonomiczną małżonków i jakością oraz stabilnością małżeństwa. Zaproponowano model, w którym trudności ekonomiczne (brak odpowiedniej w stosunku do potrzeb ilości środków finansowych, brak stałej pracy itp.) wywołują rodzaj „stresu ekonomicznego”, który powoduje napięcia w relacjach między małżonkami, obniża jakość małżeństwa i ostatecznie przy-

⁴⁶ D.A. M a d d e n - D e r d i c h, J.A. A r d i t t i, *The Ties That Bind: Attachment Between Former Spouses*, „Family Relations” 1999, nr 3.

⁴⁷ J.M. T s c h a n n, J.R. J o n h s t o n, J.S. W a l l e r s t e i n, *Resources, Stressors and Attachment as Predictors for Adult Adjustment after Divorce: A Longitudinal Study*, „Journal of Marriage and the Family” 1989, Nov.

⁴⁸ Y. L a v e e, *Linear Structural Relationship (LISREL) in Family Research*, „Journal of Marriage and the Family” 1988, Nov.

czynia się do jego destabilizacji. LISREL posłużył do empirycznej weryfikacji modelu⁴⁹.

Analiza czynnikowa – w odróżnieniu od scharakteryzowanych wyżej kilku metod analizy wielozmiennowej – jest często stosowana w naukach społecznych, zwłaszcza w psychologii. W socjologii – również w socjologicznych badaniach nad rozwodami – wykorzystywana jest stosunkowo rzadko. Metoda ta służy identyfikacji zmiennych, zwanych czynnikami wyjaśniającymi wzory korelacji występujące w ramach zbiorów obserwowanych zmiennych. Analiza czynnikowa jest często wykorzystywana w redukcji danych do identyfikacji niewielkiej liczby czynników wyjaśniających większą część wariancji z dużej liczby zmiennych. Analiza czynnikowa może być również wykorzystywana do ustalania hipotez dotyczących mechanizmów przyczynowo-skutkowych lub klasyfikowania zmiennych do dalszych analiz (na przykład do identyfikowania zależności liniowej przed rozpoczęciem analizy regresji liniowej).

Przykładem zastosowania analizy czynnikowej w badaniach nad rozwodami jest projekt badawczy, którego celem było ukazanie związku między problemami małżeńskimi, które doprowadziły do podjęcia decyzji rozwodowej i przystosowaniem porozwodowym. Analiza czynnikowa została wykorzystana do utworzenia listy grup problemów oraz do skonstruowania listy strategii przystosowania porozwodowego⁵⁰. Analizę czynnikową wykorzystano także w badaniach, których celem była identyfikacja i pomiar znaczenia najważniejszych źródeł stresu związanego z procesem rozwodowym: (1) wydarzeń związanych z rozwodem wraz z ich psychologiczną percepcją, (2) obszarem i głębokością zmian, które zaszły po rozwodzie. Analiza czynnikowa została zastosowana w celu identyfikacji „czynników stresu” w obydwu wymienionych polach analizy⁵¹.

Ideą analizy skupień (ang. *cluster analysis*) jest dzielenie zbioru obiektów w klasy bez istniejącego wcześniej kryterium zewnętrznego. Kryterium klasyfikacyjne, czyli reguła pozwalająca przypisać obiekty do poszczególnych grup, jest tworzone w trakcie analizy na podstawie uwzględnionych cech.

⁴⁹ R.D. Conger, G.H. Elder, F.O. Lorenz, K.J. Conger, R.L. Simons, L.B. Whitebeck, S. Huck, J.M. Melby, *Linking Economic Hardship to Marital Quality and Stability*, „Journal of Marriage and the Family” 1990, Aug.

⁵⁰ W.H. Berman, D.C. Turk, *Adaptation to Divorce: Problems and Coping Strategies*, „Journal of Marriage and the Family” 1981, Feb.

⁵¹ Ch. Buehler, M. Langenbrunner, *Divorce-Related Stressors: Occurrence, Disruptiveness, and Area of Life Change*, „Journal of Divorce” 1987, nr 1.

Analiza skupień wyodrębnia grupy obiektów według zasady podobieństwa. Grupy obiektów (skupienia) są tworzone przez obiekty, które są bardziej podobne do obiektów współtworzących dane skupienie niż do obiektów innych skupień. Kryterium klasyfikacyjnym jest matematycznie zdefiniowane podobieństwo. Innymi słowy – im bliżej siebie (na wykresie) położone są obiekty lub cechy, tym bardziej są do siebie podobne lub tym częściej współwystępują⁵². Ta procedura umożliwia podjęcie próby identyfikacji względnie jednorodnych grup obserwacji (lub zmiennych) na podstawie wybranej charakterystyki i z wykorzystaniem algorytmu, który rozpoczynając działanie w sytuacji, kiedy każda obserwacja (lub zmienna) jest skupiona oddzielnie, łączy skupienia aż do momentu, kiedy pozostanie tylko jedno.

Przykładem zastosowania analizy skupień w badaniach nad rozwodami jest próba budowy typologii rozpadu małżeństwa⁵³. Wyróżniono trzy typy rozpadu: szybki, stopniowy i rozciągnięty w czasie. Typy te scharakteryzowano w dwóch płaszczyznach: czasu trwania procesu rozpadu oraz kategorii przyczyn wywołujących rozpad. Analiza skupień posłużyła do grupowania podobnych „trajektorii” rozpadu⁵⁴.

UWAGI KOŃCOWE

W socjologicznych badaniach rozwodów zdecydowanie najczęściej wykorzystywana jest wielokrotna analiza regresji zarówno jako samodzielna metoda analizy, jak też w zestawieniu z innymi metodami. Stosunkowo często stosuje się również regresję logistyczną i modele probit, natomiast rzadziej analizę przeżycia. Rzadko wykorzystywana jest analiza korelacyjna, sporadycznie zaś – korelacja kanoniczna, analiza ścieżek, analiza czynnikowa, model LISREL i analiza skupień⁵⁵.

⁵² Por. T. M a r e k, Cz. N o w o r o l, *Zarys analizy skupień (niehierarchiczne i hierarchiczne techniki skupiania)*, w: B r z e z i ń s k i (red.), *Wielozmiennowe modele*, s. 185-187; M. Z a k r z e w s k a, *O miarach podobieństwa obiektów i cech przydatnych w psychologicznych zastosowaniach analizy skupień*, tamże, s. 206-210.

⁵³ J.J. P o n z e t t i, R.M. C a t e, *The Divorce Process: Toward a Typology of Marital Dissolution*, „Journal of Divorce” 1988, nr 3.

⁵⁴ Więcej na temat omawianej typologii: P. R y d z e w s k i, *Modele i typologie procesu rozwodowego*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny” 1991, nr 1.

⁵⁵ W analizach zaprezentowanych w tym artykule pominięto zastosowanie testów nieparametrycznych, takich jak np. chi-kwadrat.

Można pokusić się też o próbę prognozy: co – i jak – będzie badane w przyszłości w zakresie socjologicznych badań rozwodów? Czy czeka nas ciągła weryfikacja, w obliczu wyeksploatowanej eksploracji? Nowe pola badawcze – to obecnie duża rzadkość. Obserwując trendy w ciągu 20 lat objętych badaniami, można postawić tezę, iż *novum* będą stanowiły nie problemy, lecz raczej metody badań, w szczególności metody analizy danych. Ta tendencja uwidacznia się obecnie bardzo wyraźnie. Dostępność profesjonalnych programów do analizy statystycznej – przede wszystkim SPSS – wyraźnie inspiruje badaczy. Zawansowanie analiz statystycznych z końca okresu obserwacji jest zdecydowanie większe niż na jego początku. Z tą tendencją – niewątpliwie pozytywną – mogą wiązać się jednak pewne zagrożenia – przede wszystkim ryzyko „fetyszycyzacji” statystyki. Autor niniejszych rozważań – sam zafascynowany współczesnymi możliwościami obliczeniowymi – dostrzega jednak istotne ryzyko: sytuację, w której socjologowie wyjdą zbyt daleko poza swoją *ojkumene*. Już dziś, czytając pewne teksty, odnosi się wrażenie, iż niektórzy autorzy niezbyt pewnie poruszają się po szerokich wodach bardziej zaawansowanych matematycznie analiz.

Być może jest to bardziej ogólny trend w naukach społecznych, który można schematycznie zdefiniować w następujący sposób: ogólna eksploracja pola badań – eksploracja coraz bardziej szczegółowa – wyczerpywanie się nowych pól eksploracji – weryfikacja (zarówno dynamiczna, jak i metodologiczna). Jeżeli tak rzeczywiście jest, to można powiedzieć, iż nauki społeczne znalazły się w wyjątkowo interesującej sytuacji, gdyż dzięki komputeryzacji i lawinowemu wzrostowi mocy obliczeniowych oraz możliwościom płynącym z dostępnych programów do analizy statystycznej zarysowany wyżej schemat może być poszerzony o dodatkowy etap: eksploracji statystycznej w znaczeniu znajdowania nowych pól badań dzięki możliwościom zaawansowanych metod analizy danych.

DATA ANALYSIS METHODS
IN SOCIOLOGICAL STUDIES OF DIVORCES

S u m m a r y

The aim of the article is to give an answer to the question: What data analysis methods are most often used in field studies of divorces with the participation of divorced persons and/or members of their families? The source basis was constituted by a selection of sociological articles on divorces published in renowned American academic journals in 1980-2000. In sociological studies of divorces multiple regression analysis is decidedly most often used – both as an autonomous method of analysis and together with other methods. Also logistic regression and probit models are relatively frequently used, whereas experience analysis is used more rarely. Correlation analysis is seldom used; and canonical correlation, path analysis, factor analysis, the LISREL model and cluster analysis are used only sporadically.

Translated by Tadeusz Karłowicz

Słowa kluczowe: rozwód, metody analizy danych.

Key words: divorce, data analysis methods.