

Wstęp

Niniejsza książka jest poświęcona szeroko rozumianej analizie zależności zmiennych niemetrycznych ze szczególnym uwzględnieniem analizy logarytmiczno-liniowej. Celem monografii jest prezentacja teoretycznych i metodologicznych podstaw analizy logarytmiczno-liniowej oraz jej zastosowanie w analizie danych z wykorzystaniem programu R. W monografii przedstawiono również inne metody analizy zależności, takie jak analiza korelacji, analiza korespondencji oraz analiza klas ukrytych, a także metody wizualizacji danych zapisanych w postaci tablicy kontyngencji.

Książka ma charakter teoretyczno-empiryczny i składa się z pięciu rozdziałów.

W rozdziale pierwszym zaprezentowano podstawowe pojęcia związane z analizą zależności zmiennych niemetrycznych dla tablic dwu- i wielowymiarowych. Przedstawiono tam opisowe metody wykorzystywane do badania siły zależności w przypadku tablic o wymiarach 2×2 , do których należą: współczynnik chi-kwadrat, korekta Yatesa, statystyka Fishera, współczynnik McNemara oraz łączona statystka chi-kwadrat. W kolejnej części rozdziału zaprezentowano mierniki stosowane w przypadku tablic o wymiarach $H \times J$, do których należą współczynniki: chi-kwadrat, Cressie'a i Reada, Yule'a, Pearsona, zbieżności Czuprowa oraz Goodmana i Kruskala. W dalszej części rozdziału przedstawiono metody analizy tablic o wymiarach $2 \times 2 \times K$, do których należą: współczynnik chi-kwadrat dla łączonych tablic kontyngencji, statystyka Mantela–Haenszela, zaprezentowano także paradoks Simpsona. Następnie przedstawiono metody analizy tablic wielowymiarowych $H \times J \times K$. Rozdział ten zakończono opisem metod analizy zależności zmiennych porządkowych, do których należą: współczynniki τ i W Kendalla, γ Goodmana i Kruskala, d Somersa oraz współczynnik korelacji Spearmana.

Rozdział drugi poświęcono analizie korespondencji, której głównym celem jest ocena współwystępowania kategorii zmiennych nominalnych. Metoda ta pozwala na graficzne przedstawienie związków i relacji zachodzących pomiędzy kategoriami zmiennych. W rozdziale tym omówiono klasyczną analizę korespondencji, która jest wykorzystywana w przypadku tablic dwuwymiarowych, a także sposoby zapisu danych w przypadku wielowymiarowej analizy korespondencji w postaci tablicy Burta, złożonej macierzy znaczników, wielowymiarowej tablicy kontyngencji oraz łączonej tablicy kontyngencji. Zaprezentowano tam także rozkład macierzy według wartości osobliwych, który pozwala na wyznaczenie współrzędnych punktów reprezentujących kategorie wierszowe i kolumnowe. Porównano cztery sposo-

by rozkładu macierzy różnic standaryzowanych według wartości osobliwych, które przeprowadzono na podstawie autorskich procedur napisanych w programie R.

Rozdział trzeci poświęcono analizie logarytmiczno-liniowej, która jest modelową metodą analizy zależności zmiennych niemetrycznych. W metodzie tej modele logarytmiczno-liniowe uwzględniają interakcje pomiędzy zmiennymi i wskazują na charakter związku między nimi, dzięki czemu można opisać wiele rodzajów zależności. Celem analizy jest zbadanie, czy rozkład liczebności we wnętrzu danej tablicy kontyngencji może zostać wyrażony za pomocą prostszej struktury niż ta, która uwzględnia wszystkie dostępne zmienne i interakcje. W metodzie tej buduje się wiele modeli według zasady hierarchiczności, a wybór modelu najlepszego jest dokonywany w taki sposób, by różnica między liczebnościami empirycznymi a teoretycznymi była możliwie najmniejsza. Do oceny jakości modeli wykorzystuje się liczne kryteria określające stopień tych odchyień. Wybrany zostaje model o jak najmniejszej złożoności, który jednocześnie cechuje się wystarczającym dopasowaniem do danych.

Rozdział czwarty dotyczy analizy klas ukrytych, która znajduje zastosowanie wówczas, gdy badane zmienne są bezpośrednio nieobserwowalne. W metodzie tej zarówno zmienne obserwowalne, jak i ukryte mają charakter zmiennych skokowych. Analiza klas ukrytych ma na celu znalezienie oraz zidentyfikowanie odpowiedniej liczby klas ukrytych, w których zmienne obserwowalne są od siebie niezależne. Zbudowany model przydziela w efekcie obserwacji do klas ukrytych, a w dalszym etapie pozwala na przypuszczenie, jak zmienne obserwowalne zachowują się pod wpływem zmiennych ukrytych. W rozdziale tym zaprezentowano także liczne kryteria służące do oceny modelu.

W rozdziale piątym zaprezentowano metody wizualizacji danych zapisanych w tablicach dwu- i wielowymiarowych. Do omawianych w rozdziale wykresów należą: czteropolowy, sitkowy, mozaikowy, asocjacji oraz dwuwarstwowy. Graficzna forma prezentacji danych w postaci zaawansowanych wykresów pełni istotną rolę w procesie analizy, a niekiedy zastępuje nawet formalne kryteria wykorzystywane w celu wyboru modelu najlepszego.

Niniejsza monografia jest przeznaczona dla pracowników naukowych zajmujących się zastosowaniem metod analizy danych jakościowych w naukach społecznych, a w szczególności w psychologii, socjologii, ekonomii czy politologii. Ponadto odbiorcami książki mogą być studenci uczelni wyższych studiujący na specjalnościach ilościowych (informatyka ekonomiczna, statystyka i ekonometria, analityka gospodarcza), doktoranci, którzy wykorzystują w swych badaniach analizę danych jakościowych oraz analitycy rynku i pracownicy agencji marketingowych przeprowadzający analizy danych niemetrycznych z wykorzystaniem programu R.

W badaniach empirycznych i we wszystkich obliczeniach prezentowanych w książce wykorzystano program R. Wersję instalacyjną programu oraz dodatkowe pakiety zaprezentowane w niniejszej monografii można pobrać ze strony pod adresem: <http://www.r-project.org/>.

Pragnę serdecznie podziękować Panu Profesorowi Eugeniuszowi Gatnarowi za życzliwą pomoc, cenne uwagi merytoryczne oraz opiekę naukową, na którą zawsze mogłam liczyć. Dziękuję również Panu Profesorowi Markowi Walesiakowi z Uniwersytetu Ekonomicznego we Wrocławiu oraz Panu Profesorowi Adamowi Saganowi z Uniwersytetu Ekonomicznego w Krakowie za cenne i konstruktywne uwagi, które wpłynęły na ostateczny kształt monografii.

Tę książkę pragnę zadedykować moim Rodzicom, którzy są *spiritus movens* tego, co w moim życiu najwspanialsze.