

Analiza
danych jakościowych
i symbolicznych
z wykorzystaniem programu R

Redakcja naukowa

Eugeniusz Gatnar

Marek Walesiak

Wydawnictwo C.H. Beck 

Spis treści

Wstęp	9
Rozdział 1. Wprowadzenie do analizy danych jakościowych i symbolicznych	13
1.1. Macierz danych i tablica danych	13
1.2. Miary odległości	17
1.2.1. Dane porządkowe	17
1.2.2. Dane symboliczne	18
1.3. Dyskretyzacja zmiennych ilościowych	25
1.4. Wybrane rozkłady prawdopodobieństwa zmiennych dyskretnych	35
1.5. Wizualizacja danych	39
Rozdział 2. Analiza korespondencji	52
2.1. Wprowadzenie	52
2.2. Tablice kontyngencji	52
2.3. Analiza zależności między zmiennymi	53
2.4. Analiza korespondencji dwu i wielu zmiennych	57
2.5. Wizualizacja wyników klasycznej i wielowymiarowej analizy korespondencji	67
2.6. Zastosowania z wykorzystaniem programu R	68
Rozdział 3. Modele logarytmiczno-liniowe	81
3.1. Wprowadzenie	81
3.2. Klasyczny model logarytmiczno-liniowy	82
3.3. Hierarchiczne modele logarytmiczno-liniowe	88
3.4. Miary dopasowania modeli logarytmiczno-liniowych	90
3.5. Zastosowania z wykorzystaniem programu R	92
Rozdział 4. Modelowanie i prognozowanie zmiennych dwumianowych	99
4.1. Wprowadzenie	99
4.2. Liniowy model prawdopodobieństwa (LMP)	99
4.3. Modele logitowe i probitowe	101
4.4. Prognozy na podstawie modeli dwumianowych	103
4.5. Zastosowania z wykorzystaniem programu R	104
Rozdział 5. Modelowanie zmiennych wielomianowych	112
5.1. Wprowadzenie	112
5.2. Wielomianowy model logitowy	113
5.3. Warunkowy model logitowy	114
5.4. Analiza historii zdarzeń	115
5.5. Zastosowania z wykorzystaniem programu R	116

Rozdział 6. Analiza wariancji	131
6.1. Podstawy teoretyczne	131
6.1.1. Jednoczynnikowa analiza wariancji	132
6.1.2. Dwuczynnikowa analiza wariancji	134
6.1.3. Założenia analizy wariancji	139
6.1.4. Testy <i>post hoc</i>	142
6.2. Podstawowe schematy badań	143
6.3. Zastosowania z wykorzystaniem programu R	147
Rozdział 7. Analiza skupień i porządkowanie liniowe na podstawie danych porządkowych	165
7.1. Wprowadzenie	165
7.2. Analiza skupień na podstawie danych porządkowych	165
7.3. Porządkowanie liniowe na podstawie danych porządkowych	170
7.4. Zastosowania z wykorzystaniem programu R	172
Rozdział 8. Drzewa klasyfikacyjne i regresyjne dla jakościowych zmiennych objaśniających	181
8.1. Podstawy teoretyczne	181
8.2. Drzewa klasyfikacyjne i regresyjne	185
8.3. Dobór jakościowych zmiennych objaśniających	188
8.4. Określenie optymalnej postaci modelu	190
8.5. Zastosowania z wykorzystaniem programu R	192
Rozdział 9. Modele klas ukrytych dla danych jakościowych	204
9.1. Wprowadzenie	204
9.2. Model klas ukrytych dla zmiennych binarnych i wielomianowych	204
9.3. Model regresji klas ukrytych	206
9.4. Zastosowania z wykorzystaniem programu R	207
Rozdział 10. Modele mieszanek dla danych jakościowych	223
10.1. Wprowadzenie	223
10.2. Model GLM	223
10.3. Modele mieszanek – podstawy teoretyczne	224
10.4. Modele mieszanek rozkładów dwumianowych	227
10.5. Modele mieszanek rozkładów Poissona	229
10.6. Zastosowania z wykorzystaniem programu R	232
Rozdział 11. Skalowanie wielowymiarowe na podstawie danych jakościowych i symbolicznych	242
11.1. Procedury skalowania wielowymiarowego na podstawie danych jakościowych	242
11.2. Analiza <i>unfolding</i>	246
11.3. Skalowanie wielowymiarowe na podstawie danych symbolicznych	248
11.4. Zastosowania z wykorzystaniem programu R	255
Rozdział 12. Analiza skupień na podstawie danych symbolicznych	262
12.1. Wprowadzenie	262
12.2. Podejścia i metody klasyfikacji danych symbolicznych	262
12.3. Procedura klasyfikacji danych symbolicznych	264
12.4. Zastosowania z wykorzystaniem programu R	269

Rozdział 13. Analiza dyskryminacyjna i drzewa klasyfikacyjne na podstawie danych symbolicznych	280
13.1. Analiza dyskryminacyjna bazująca na estymatorach intensywności	280
13.2. Drzewa klasyfikacyjne bazujące na optymalnym podziale	282
13.3. Bayesowskie drzewa klasyfikacyjne	285
13.4. Zastosowania z wykorzystaniem programu R	286
Dodatek A. Format danych symbolicznych	292
Bibliografia	294
Indeks	304