

Spis treści

Przedmowa	17
MONIKA KSIĄŻEK	
Rozdział I	
Analiza danych jakościowych	25
I. Teoria	25
I.1. Wprowadzenie	25
I.2. Jednowymiarowa analiza zmiennych jakościowych	26
I.3. Analiza zależności zmiennych jakościowych	31
I.3.1. Test równości proporcji	32
I.3.2. Miary zależności	33
I.3.3. Testy niezależności	38
I.3.4. Testy i mierniki dla zmiennych porządkowych	40
I.3.5. Analiza zależności dwóch zmiennych jakościowych w warstwach wyznaczanych przez inne zmienne jakościowe	42
I.4. Modele log-liniowe	45
I.5. Zmienne jakościowe jako zmienne objaśniające	47
I.5.1. Kodowanie zmiennych jakościowych	47
I.5.2. Problemy związane z obecnością zmiennych jakościowych w modelu ..	50
I.5.3. Interakcje zmiennych jakościowych	52
I.5.4. Korzyści ze zmiany skali pomiaru zmiennych	53
I.6. Zmienne jakościowe jako zmienne objaśniane	54
I.6.1. Uogólnione modele liniowe dla zmiennych jakościowych	54
I.6.2. Interpretacja parametrów	59
I.6.3. Badanie istotności statystycznej parametrów	60
I.6.4. Ocena jakości dopasowania modelu	61
II. Przykłady	63
II.1. Analiza jednowymiarowa	68
II.2. Analiza tabeli 2×2	71
II.3. Analiza tabeli 2×2 w podgrupach	77
II.4. Model log-liniowy	86

II.5. Kodowanie	93
II.6. Binarna regresja logistyczna	101
II.7. Wielomianowa i porządkowa regresja logistyczna	111
II.8. Uwzględnianie wpływów nieliniowych	120
II.9. Interakcje	130
Bibliografia	137

IGA SIKORSKA

Rozdział II

Analiza zmiennych ukrytych	139
I. Model klas ukrytych (LCA)	139
I.1. Wprowadzenie	139
I.2. Zapis modelu	144
I.2.1. Założenia modelu klas ukrytych	147
I.2.2. Parametry modelu klas ukrytych	148
I.3. Estymacja modelu klas ukrytych	149
I.3.1. Estymacja parametrów w SAS	153
I.4. Weryfikacja modelu klas ukrytych	154
I.4.1. Braki danych	161
I.4.2. Ograniczenia nakładane na parametry	162
I.4.3. Liczba klas ukrytych	163
I.4.4. Interpretacja klas ukrytych	165
I.4.5. Homogeniczność i rozróżnialność modeli klas ukrytych	166
I.5. Zmienne grupujące w modelu klas ukrytych	168
I.6. Zmienne kontrolne w modelu klas ukrytych	171
I.6.1. Weryfikacja modelu ze zmiennymi kontrolnymi	172
I.7. Procedury LCA i LTA	173
I.8. Przykład modelu klas ukrytych	176
II. Model stanów ukrytych (LTA)	193
II.1. Wprowadzenie	193
II.2. Zapis modelu	193
II.3. Estymacja i weryfikacja modelu stanów ukrytych	198
II.3.1. Braki danych	199
II.3.2. Ograniczenia nakładane na parametry	199
II.4. Zmienne grupujące w modelu stanów ukrytych	202
II.5. Zmienne kontrolne w modelu stanów ukrytych	204
II.6. Przykład modelu stanów ukrytych	206
Bibliografia	220

EWA FRĄTCZAK, MAŁGORZATA MIANOWSKA

Rozdział III

Modele mieszane	223
I. Podstawy teoretyczne	223
I.1. Liniowy model mieszany	228
I.1.1. Wprowadzenie	228
I.1.2. Zapis liniowego modelu mieszanego	232
I.1.3. PROC GLM i PROC MIXED	238
I.1.4. PROC HPMIXED	243
I.1.5. Diagnostyka i strategie budowy modelu	244
I.2. Uogólniony model mieszany	247
I.2.1. Wprowadzenie	247
I.2.2. Zapis uogólnionego liniowego modelu mieszanego	251
I.2.3. Procedura GLIMMIX i metody estymacji	253
I.3. Nieliniowy model mieszany	254
I.3.1. Wprowadzenie	254
I.3.2. Zapis nieliniowego modelu mieszanego	254
I.3.3. Procedura NLMIXED i metoda estymacji	255
I.4. Podsumowanie	257
II. Przykłady estymacji modeli mieszanych	259
Wprowadzenie	259
II.1. Przykład 1 – model liniowy i liniowy model mieszany	262
II.2. Przykład 2 – liniowy model mieszany z interakcją	301
II.3. Przykład 3 – model hierarchiczny	318
II.4. Przykład 4 – uogólniony liniowy model mieszany i model nieliniowy	351
II.5. Przykład 5 – estymacja modelu mieszanego w Enterprise Guide	369
Bibliografia	378
Załącznik 1. Teoria liniowych modeli mieszanych	381
1. Wprowadzenie	381
2. Zapis macierzowy	381
3. Określenie postaci modeli mieszanych	383
3.1. Ogólna postać liniowego modelu mieszanego	383
3.2. Rozkłady warunkowe i brzegowe	384
3.3. Przykład: Krzywa wzrostu z symetryczną strukturą kowariancji	386
3.4. Przykład: Układ podzielonych poletek (<i>Split-Plot Design</i>)	388
4. Estymacja parametrów, predykcja efektów losowych	392
4.1. Estymacja β i prognoza u : równania modelu mieszanego	392
4.2. Efekty losowe, grzbietowe oraz kurczenie	394

4.3. Wszystko o metodzie SWEEP	396
4.4. Największa wiarygodność i ograniczona największa wiarygodność dla parametrów kowariancji	399
5. Własności statystyczne	405
6. Wybór postaci modelu	407
6.1. Porównania modeli z wykorzystaniem testów ilorazu wiarygodności	408
6.2. Porównania modeli z wykorzystaniem kryteriów informacyjnych	409
7. Wnioskowanie i statystyki testujące	411
7.1. Wnioskowanie o parametrach kowariancji	411
7.2. Wnioskowanie o efektach stałych i losowych	412
8. Prace cytowane w załączniku	414

WIOLETTA GRZENDA

Rozdział IV

Wybrane zagadnienia estymacji bayesowskiej	419
I. Elementy teorii statystyki bayesowskiej	420
I.1. Metody bayesowskie	420
I.1.1. Twierdzenie Bayesa	422
I.1.2. Rozkłady <i>a priori</i>	427
I.1.3. Wnioskowanie bayesowskie	432
I.1.4. Uwagi ogólne dotyczące metod bayesowskich	435
I.2. Metody Monte Carlo oparte na łańcuchach Markowa	436
I.2.1. Wybrane własności łańcuchów Markowa	437
I.2.2. Algorytm Metropolis i algorytm Metropolis–Hastingsa	442
I.2.3. Próbnik Gibbsa	444
I.2.4. Algorytm próbkowania adaptacyjnego z odrzucaniem	445
I.2.5. Zagadnienia dotyczące wyboru realizacji łańcucha Markowa	445
I.2.6. Ocena zbieżności łańcuchów Markowa	447
I.2.7. Testy zbieżności łańcuchów Markowa	451
II. Przykłady zastosowań	459
II.1. Materiał empiryczny	460
II.2. Model regresji Poissona w ujęciu bayesowskim	461
II.3. Bayesowska estymacja uogólnionych modeli liniowych w systemie SAS ...	462
II.4. Przykłady bayesowskiej estymacji modeli regresji Poissona	466
II.4.1. Model Poissona z nieinformacyjnymi rozkładami normalnymi <i>a priori</i>	467
II.4.2. Model Poissona z informacyjnym rozkładem normalnym <i>a priori</i> i nieinformacyjnymi rozkładami normalnymi <i>a priori</i>	490
Bibliografia	499

KAMIL KONIKIEWICZ

Rozdział V

Data Mining	503
1. Wprowadzenie do aplikacji <i>SAS Enterprise Miner</i>	505
2. Opis danych	512
3. Podział danych	514
4. Eksploracja danych	517
5. Drzewa decyzyjne	524
5.1. Postać modelu	524
5.2. Budowa modelu	525
5.3. Dobór zmiennych i przygotowanie danych	530
5.4. Lasy losowe	540
6. Regresja logistyczna	541
6.1. Postać modelu	541
6.2. Przygotowanie danych	542
7. Sieci neuronowe MLP	554
7.1. Postać modelu	554
7.2. Uczenie sieci	556
8. Ocena i wybór modelu	562
8.1. Statystyki dopasowania	562
8.2. Przeglądanie	567
8.3. Scoring	570
Bibliografia	570

EWA FALKIEWICZ-SZPORA, ŁUKASZ LESZEWSKI

Rozdział VI

Wybrane zagadnienia jakości danych	573
I. Podstawowe pojęcia	574
I.1. Cechy dobrej jakości danych	574
I.2. Źródła złej jakości danych	577
I.3. Etapy tworzenia i transformacji informacji	579
II. Metodologia zarządzania jakością danych	580
III. Filary zarządzania jakością danych	584
III.1. Tworzenie otoczenia sprzyjającego jakości danych	585
III.2. Rozwiązania organizacyjne	586
III.3. Zapewnienie standardów w organizacji	587
III.4. Monitorowanie i mierzenie jakości danych	588
III.5. Rola hurtowni danych	591
III.6. Technologia i narzędzia	591

IV. Etapy procesu czyszczenia danych	593
IV.1. Profilowanie	594
IV.2. Czyszczenie danych	595
IV.3. Integracja danych	596
IV.4. Wzbogacanie danych	597
IV.5. Monitorowanie danych	597
V. Narzędzia i techniki jakości danych	598
V.1. DataFlux dfPower Studio	599
V.2. DataFlux Integration Server	602
V.3. SAS Data Integration Studio	603
VI. Standaryzacja danych	604
VI.1. Tworzenie schematów standaryzacyjnych	606
VI.2. Definicje standaryzacyjne	609
VII. Przykład zastosowania – implementacja procesów czyszczenia danych	610
VII.1. Profilowanie	611
VII.2. Standaryzacja	614
VII.3. Integracja danych	615
VIII. Podsumowanie	617
Bibliografia	618
Streszczenia – Abstracts	619
Biogramy – Biograms	629