

Contents

Part I Cluster Models

1	Data Mining for Visualization of Health Processes	
	(150 Patients with Pneumonia)	3
1.1	General Purpose	3
1.2	Primary Scientific Question	3
1.3	Example	3
1.4	Knime Data Miner	5
1.5	Knime Workflow	6
1.6	Box and Whiskers Plots	7
1.7	Lift Chart	7
1.8	Histogram	8
1.9	Line Plot	9
1.10	Matrix of Scatter Plots	10
1.11	Parallel Coordinates	11
1.12	Hierarchical Cluster Analysis with SOTA	
	(Self Organizing Tree Algorithm)	12
1.13	Conclusion	13
2	Training Decision Trees for a More Meaningful Accuracy	
	(150 Patients with Pneumonia)	15
2.1	General Purpose	15
2.2	Primary Scientific Question	15
2.3	Example	16
2.4	Downloading the Knime Data Miner	17
2.5	Knime Workflow	18
2.6	Conclusion	20

Part II Linear Models

3	Variance Components for Assessing the Magnitude of Random Effects (40 Patients with Paroxysmal Tachycardias)	23
3.1	General Purpose	23
3.2	Primary Scientific Question	23
3.3	Example	23
3.4	Conclusion	27
4	Ordinal Scaling for Clinical Scores with Inconsistent Intervals (900 Patients with Different Levels of Health).	29
4.1	General Purpose	29
4.2	Primary Scientific Questions	29
4.3	Example	29
4.4	Conclusion	33
5	Loglinear Models for Assessing Incident Rates with Varying Incident Risks (12 Populations with Different Drink Consumption Patterns)	35
5.1	General Purpose	35
5.2	Primary Scientific Question	36
5.3	Example	36
5.4	Conclusion	38
6	Loglinear Modeling for Outcome Categories (Quality of Life Scores in 445 Patients with Different Personal Characteristics).	39
6.1	General Purpose	39
6.2	Primary Scientific Question	39
6.3	Example	39
6.4	Conclusion	43
7	Heterogeneity in Clinical Research: Mechanisms Responsible	45
7.1	General Purpose	45
7.2	Primary Scientific Question	45
7.3	Example	45
7.4	Conclusion	48
8	Performance Evaluation of Novel Diagnostic Tests (650 and 588 Patients with Peripheral Vascular Disease)	49
8.1	General Purpose	49
8.2	Primary Scientific Question	49

8.3	Example	49
8.3.1	Binary Logistic Regression.	52
8.3.2	c-Statistics	53
8.4	Conclusion.	55

Part III Rules Models

9	Spectral Plots for High Sensitivity Assessment of Periodicity (6 Years' Monthly C Reactive Protein Levels)	59
9.1	General Purpose	59
9.2	Specific Scientific Question	59
9.3	Example	59
9.4	Conclusion.	63
10	Runs Test for Identifying Best Regression Models (21 Estimates of Quantity and Quality of Patient Care)	65
10.1	General Purpose	65
10.2	Primary Scientific Question	65
10.3	Example	65
10.4	Conclusion.	69
11	Evolutionary Operations for Process Improvement (8 Operation Room Air Condition Settings)	71
11.1	General Purpose	71
11.2	Specific Scientific Question	71
11.3	Example	72
11.4	Conclusion.	74
12	Bayesian Networks for Cause Effect Modeling (600 Patients Assessed for Longevity Factors)	75
12.1	General Purpose	75
12.2	Primary Scientific Question	75
12.3	Example	76
12.4	Binary Logistic Regression in SPSS	76
12.5	Konstanz Information Miner (Knime)	77
12.6	Knime Workflow	78
12.6.1	Confusion Matrix	79
12.6.2	Accuracy Statistics	79
12.7	Conclusion.	80

13	Support Vector Machines for Imperfect Nonlinear Data (200 Patients with Sepsis)	81
13.1	General Purpose	81
13.2	Primary Scientific Question	81
13.3	Example	82
13.4	Knime Data Miner	82
13.5	Knime Workflow	83
13.6	File Reader Node	83
13.7	The Nodes X-Partitioner, SVM Learner, SVM Predictor, X-Aggregator	84
13.8	Error Rates	84
13.9	Prediction Table	85
13.10	Conclusion	85
14	Multiple Response Sets for Visualizing Clinical Data Trends (811 Patient Visits to General Practitioners)	87
14.1	General Purpose	87
14.2	Specific Scientific Question	87
14.3	Example	87
14.4	Conclusion	93
15	Protein and DNA Sequence Mining	95
15.1	General Purpose	95
15.2	Specific Scientific Question	96
15.3	Data Base Systems on the Internet	96
15.4	Example 1	97
15.5	Example 2	98
15.6	Example 3	98
15.7	Example 4	99
15.8	Conclusion	100
16	Iteration Methods for Crossvalidations (150 Patients with Pneumonia)	101
16.1	General Purpose	101
16.2	Primary Scientific Question	101
16.3	Example	101
16.4	Downloading the Knime Data Miner	102
16.5	Knime Workflow	103
16.6	Crossvalidation	104
16.7	Conclusion	105

17	Testing Parallel-Groups with Different Sample Sizes and Variances (5 Parallel-Group Studies).	107
17.1	General Purpose	107
17.2	Primary Scientific Question	107
17.3	Examples	108
17.4	Conclusion	109
18	Association Rules Between Exposure and Outcome (50 and 60 Patients with Coronary Risk Factors)	111
18.1	General Purpose	111
18.2	Primary Scientific Question	111
18.3	Example	111
	18.3.1 Example One	113
	18.3.2 Example Two	114
18.4	Conclusion	115
19	Confidence Intervals for Proportions and Differences in Proportions	117
19.1	General Purpose	117
19.2	Primary Scientific Question	117
19.3	Example	118
	19.3.1 Confidence Intervals of Proportions	118
	19.3.2 Confidence Intervals of Differences in Proportions	119
19.4	Conclusion	120
20	Ratio Statistics for Efficacy Analysis of New Drugs (50 Patients Treated for Hypercholesterolemia)	121
20.1	General Purpose	121
20.2	Primary Scientific Question	121
20.3	Example	122
20.4	Conclusion	125
	Index	127

Machine Learning in Medicine - Cookbook Three

Cleophas, T.J.; Zwinderman, A.H.

2014, XIII, 131 p. 37 illus., Softcover

ISBN: 978-3-319-12162-8