

Contents

Part I Cluster Models

1	Hierarchical Clustering and K-means Clustering to Identify Subgroups in Surveys (50 Patients)	3
	General Purpose	3
	Specific Scientific Question	3
	Hierarchical Cluster Analysis	4
	K-means Cluster Analysis	6
	Conclusion	7
	Note	8
2	Density-Based Clustering to Identify Outlier Groups in Otherwise Homogeneous Data (50 Patients)	9
	General Purpose	9
	Specific Scientific Question	9
	Density-Based Cluster Analysis	10
	Conclusion	11
	Note	11
3	Two Step Clustering to Identify Subgroups and Predict Subgroup Memberships in Individual Future Patients (120 Patients)	13
	General Purpose	13
	Specific Scientific Question	13
	The Computer Teaches Itself to Make Predictions	14
	Conclusion	15
	Note	15

Part II Linear Models

4	Linear, Logistic, and Cox Regression for Outcome Prediction with Unpaired Data (20, 55, and 60 Patients)	19
	General Purpose	19

Specific Scientific Question	19
Linear Regression, the Computer Teaches Itself to Make Predictions	19
Conclusion	21
Note	21
Logistic Regression, the Computer Teaches Itself to Make Predictions	22
Conclusion	24
Note	24
Cox Regression, the Computer Teaches Itself to Make Predictions	24
Conclusion	26
Note	26
5 Generalized Linear Models for Outcome Prediction with Paired Data (100 Patients and 139 Physicians)	29
General Purpose.	29
Specific Scientific Question	29
Generalized Linear Modeling, the Computer Teaches Itself to Make Predictions	29
Conclusion	31
Generalized Estimation Equations, the Computer Teaches Itself to Make Predictions	32
Conclusion	34
Note	35
6 Generalized Linear Models for Predicting Event-Rates (50 Patients)	37
General Purpose.	37
Specific Scientific Question	37
The Computer Teaches Itself to Make Predictions	38
Conclusion	40
Note	41
7 Factor Analysis and Partial Least Squares for Complex-Data Reduction (250 Patients)	43
General Purpose.	43
Specific Scientific Question	43
Factor Analysis	44
Partial Least Squares Analysis.	46
Traditional Linear Regression	48
Conclusion	48
Note	49

8	Optimal Scaling of High-Sensitivity Analysis of Health Predictors (250 Patients)	51
	General Purpose	51
	Specific Scientific Question	51
	Traditional Multiple Linear Regression	52
	Optimal Scaling Without Regularization	53
	Optimal Scaling with Ridge Regression	54
	Optimal Scaling with Lasso Regression	54
	Optimal Scaling with Elastic Net Regression	55
	Conclusion	56
	Note	56
9	Discriminant Analysis for Making a Diagnosis from Multiple Outcomes (45 Patients)	57
	General Purpose	57
	Specific Scientific Question	57
	The Computer Teaches Itself to Make Predictions	58
	Conclusion	61
	Note	61
10	Weighted Least Squares for Adjusting Efficacy Data with Inconsistent Spread (78 Patients)	63
	General Purpose	63
	Specific Scientific Question	63
	Weighted Least Squares	64
	Conclusion	66
	Note	66
11	Partial Correlations for Removing Interaction Effects from Efficacy Data (64 Patients)	67
	General Purpose	67
	Specific Scientific Question	67
	Partial Correlations	68
	Conclusion	70
	Note	71
12	Canonical Regression for Overall Statistics of Multivariate Data (250 Patients)	73
	General Purpose	73
	Specific Scientific Question	73
	Canonical Regression	74
	Conclusion	76
	Note	77

Part III Rules Models

13 Neural Networks for Assessing Relationships that are Typically Nonlinear (90 Patients)	81
General Purpose.	81
Specific Scientific Question	81
The Computer Teaches Itself to Make Predictions	82
Conclusion	83
Note.	83
14 Complex Samples Methodologies for Unbiased Sampling (9,678 Persons)	85
General Purpose.	85
Specific Scientific Question	85
The Computer Teaches Itself to Predict Current Health Scores from Previous Health Scores	87
The Computer Teaches Itself to Predict Odds Ratios of Current Health Scores Versus Previous Health Scores	88
Conclusion	90
Note.	90
15 Correspondence Analysis for Identifying the Best of Multiple Treatments in Multiple Groups (217 Patients)	91
General Purpose.	91
Specific Scientific Question	91
Correspondence Analysis	92
Conclusion	95
Note.	95
16 Decision Trees for Decision Analysis (1,004 and 953 Patients)	97
General Purpose.	97
Specific Scientific Question	97
Decision Trees with a Binary Outcome	97
Decision Trees with a Continuous Outcome	101
Conclusion	104
Note.	104
17 Multidimensional Scaling for Visualizing Experienced Drug Efficacies (14 Pain-Killers and 42 Patients)	105
General Purpose.	105
Specific Scientific Question	105
Proximity Scaling	105
Preference Scaling	108
Conclusion	112
Note.	113

18 Stochastic Processes for Long Term Predictions from Short Term Observations	115
General Purpose.	115
Specific Scientific Questions	115
Conclusion	120
Note	121
19 Optimal Binning for Finding High Risk Cut-offs (1,445 Families)	123
General Purpose.	123
Specific Scientific Question	123
Optimal Binning	124
Conclusion	127
Note	127
20 Conjoint Analysis for Determining the Most Appreciated Properties of Medicines to be Developed (15 Physicians)	129
General Purpose.	129
Specific Scientific Question	129
Constructing an Analysis Plan	129
Performing the Final Analysis	131
Conclusion	134
Note	134
Index	135

<http://www.springer.com/978-3-319-04180-3>

Machine Learning in Medicine - Cookbook

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

2014, XI, 137 p. 14 illus., Softcover

ISBN: 978-3-319-04180-3