# Contents

**Preface** . . . . . xi

**Introduction** . . . . . xv
Gilbert SAPORTA

**Part 1. Clustering and Regression** . . . . . 1

**Chapter 1. Cluster Validation by Measurement of Clustering Characteristics Relevant to the User** . . . . . 3
Christian HENNIG

1.1. Introduction. . . . . 3
1.2. General notation . . . . . 5
1.3. Aspects of cluster validity . . . . . 6
1.3.1. Small within-cluster dissimilarities . . . . . 6
1.3.2. Between-cluster separation . . . . . 7
1.3.3. Representation of objects by centroids . . . . . 7
1.3.4. Representation of dissimilarity structure by clustering . . . . . 8
1.3.5. Small within-cluster gaps . . . . . 9
1.3.6. Density modes and valleys . . . . . 9
1.3.7. Uniform within-cluster density . . . . . 12
1.3.8. Entropy . . . . . 12
1.3.9. Parsimony . . . . . 13
1.3.10. Similarity to homogeneous distributional shapes . . . . . 13
1.3.11. Stability . . . . . 13
1.3.12. Further Aspects . . . . . 14
1.4. Aggregation of indexes . . . . . 14
1.5. Random clusterings for calibrating indexes . . . . . 15
1.5.1. Stupid K-centroids clustering . . . . . 16

1.5.2. Stupid nearest neighbors clustering . . . . . 16
1.5.3. Calibration . . . . . 17
1.6. Examples . . . . . 18
1.6.1. Artificial data set . . . . . 18
1.6.2. Tetragonula bees data . . . . . 20
1.7. Conclusion . . . . . 22
1.8. Acknowledgment . . . . . 23
1.9. References . . . . . 23

**Chapter 2. Histogram-Based Clustering of Sensor Network Data** . . 25
Antonio BALZANELLA and Rosanna VERDE

2.1. Introduction. . . . . 25
2.2. Time series data stream clustering . . . . . 28
2.2.1. Local clustering of histogram data . . . . . 30
2.2.2. Online proximity matrix updating . . . . . 32
2.2.3. Off-line partitioning through the dynamic clustering algorithm for dissimilarity tables . . . . . 33
2.3. Results on real data . . . . . 34
2.4. Conclusions. . . . . 36
2.5. References . . . . . 36

**Chapter 3. The Flexible Beta Regression Model** . . . . . 39
Sonia MIGLIORATI, Agnese M. DI BRISCO and Andrea ONGARO

3.1. Introduction. . . . . 39
3.2. The FB distribution . . . . . 41
3.2.1. The beta distribution . . . . . 41
3.2.2. The FB distribution. . . . . 41
3.2.3. Reparameterization of the FB. . . . . 42
3.3. The FB regression model. . . . . 43
3.4. Bayesian inference . . . . . 44
3.5. Illustrative application . . . . . 47
3.6. Conclusion . . . . . 48
3.7. References . . . . . 50

**Chapter 4. *S*-weighted Instrumental Variables** . . . . . 53
Jan Ámos VÍŠEK

4.1. Summarizing the previous relevant results . . . . . 53
4.2. The notations, framework, conditions and main tool . . . . . 55
4.3. *S*-weighted estimator and its consistency . . . . . 57
4.4. *S*-weighted instrumental variables and their consistency . . . . . 59
4.5. Patterns of results of simulations . . . . . 64
4.5.1. Generating the data . . . . . 65
4.5.2. Reporting the results . . . . . 66

4.6. Acknowledgment . . . . . 69
4.7. References . . . . . 69

**Part 2. Models and Modeling** . . . . . 73

**Chapter 5. Grouping Property and Decomposition of Explained Variance in Linear Regression** . . . . . 75
Henri WALLARD

5.1. Introduction . . . . . 75
5.2. CAR scores . . . . . 76
5.2.1. Definition and estimators . . . . . 76
5.2.2. Historical criticism of the CAR scores . . . . . 79
5.3. Variance decomposition methods and SVD . . . . . 79
5.4. Grouping property of variance decomposition methods . . . . . 80
5.4.1. Analysis of grouping property for CAR scores . . . . . 81
5.4.2. Demonstration with two predictors . . . . . 82
5.4.3. Analysis of grouping property using SVD . . . . . 83
5.4.4. Application to the diabetes data set . . . . . 86
5.5. Conclusions . . . . . 87
5.6. References . . . . . 88

**Chapter 6. On GARCH Models with Temporary Structural Changes** 91
Norio WATANABE and Fumiaki OKIHARA

6.1. Introduction . . . . . 91
6.2. The model . . . . . 92
6.2.1. Trend model . . . . . 92
6.2.2. Intervention GARCH model . . . . . 93
6.3. Identification . . . . . 96
6.4. Simulation . . . . . 96
6.4.1. Simulation on trend model . . . . . 96
6.4.2. Simulation on intervention trend model . . . . . 98
6.5. Application . . . . . 98
6.6. Concluding remarks . . . . . 102
6.7. References . . . . . 103

**Chapter 7. A Note on the Linear Approximation of TAR Models** . . . 105
Francesco GIORDANO, Marcella NIGLIO and Cosimo Damiano VITALE

7.1. Introduction . . . . . 105
7.2. Linear representations and linear approximations of nonlinear models . . . . . 107
7.3. Linear approximation of the TAR model . . . . . 109
7.4. References . . . . . 116

**Chapter 8. An Approximation of Social Well-Being Evaluation Using Structural Equation Modeling** . . . . . 117
Leonel SANTOS-BARRIOS, Monica RUIZ-TORRES, William GÓMEZ-DEMETRIO, Ernesto SÁNCHEZ-VERA, Ana LORGA DA SILVA and Francisco MARTÍNEZ-CASTAÑEDA

8.1. Introduction. . . . . 117
8.2. Wellness. . . . . 118
8.3. Social welfare . . . . . 118
8.4. Methodology . . . . . 119
8.5. Results. . . . . 120
8.6. Discussion . . . . . 123
8.7. Conclusions. . . . . 123
8.8. References . . . . . 123

**Chapter 9. An SEM Approach to Modeling Housing Values** . . . . . 125
Jim FREEMAN and Xin ZHAO

9.1. Introduction. . . . . 125
9.2. Data . . . . . 126
9.3. Analysis . . . . . 127
9.4. Conclusions. . . . . 134
9.5. References . . . . . 135

**Chapter 10. Evaluation of Stopping Criteria for Ranks in Solving Linear Systems** . . . . . 137
Benard ABOLA, Pitos BIGANDA, Christopher ENGSTRÖM and Sergei SILVESTROV

10.1. Introduction . . . . . 137
10.2. Methods . . . . . 139
10.2.1. Preliminaries . . . . . 139
10.2.2. Iterative methods . . . . . 140
10.3. Formulation of linear systems . . . . . 142
10.4. Stopping criteria . . . . . 143
10.5. Numerical experimentation of stopping criteria. . . . . 146
10.5.1. Convergence of stopping criterion . . . . . 147
10.5.2. Quantiles . . . . . 147
10.5.3. Kendall correlation coefficient as stopping criterion . . . . . 148
10.6. Conclusions . . . . . 150
10.7. Acknowledgments . . . . . 151
10.8. References. . . . . 151

**Chapter 11. Estimation of a Two-Variable Second-Degree Polynomial via Sampling** . . . . . 153
Ioanna PAPATSOUMA, Nikolaos FARMAKIS and Eleni KETZAKI

11.1. Introduction . . . . . 153
11.2. Proposed method . . . . . 154

11.2.1. First restriction . . . 154
11.2.2. Second restriction . . . 155
11.2.3. Third restriction . . . 156
11.2.4. Fourth restriction . . . 156
11.2.5. Fifth restriction . . . 157
11.2.6. Coefficient estimates . . . 158
11.3. Experimental approaches . . . 159
11.3.1. Experiment A . . . 159
11.3.2. Experiment B . . . 161
11.4. Conclusions . . . 163
11.5. References . . . 163

**Part 3. Estimators, Forecasting and Data Mining** . . . 165

**Chapter 12. Displaying Empirical Distributions of Conditional Quantile Estimates: An Application of Symbolic Data Analysis to the Cost Allocation Problem in Agriculture** . . . 167
Dominique DESBOIS

12.1. Conceptual framework and methodological aspects of cost allocation . . . 167
12.2. The empirical model of specific production cost estimates . . . 168
12.3. The conditional quantile estimation . . . 169
12.4. Symbolic analyses of the empirical distributions of specific costs . . . 170
12.5. The visualization and the analysis of econometric results . . . 172
12.6. Conclusion . . . 178
12.7. Acknowledgments . . . 179
12.8. References . . . 179

**Chapter 13. Frost Prediction in Apple Orchards Based upon Time Series Models** . . . 181
Monika A. TOMKOWICZ and Armin O. SCHMITT

13.1. Introduction . . . 181
13.2. Weather database . . . 182
13.3. ARIMA forecast model . . . 183
13.3.1. Stationarity and differencing . . . 184
13.3.2. Non-seasonal ARIMA models . . . 186
13.4. Model building . . . 188
13.4.1. ARIMA and LR models . . . 188
13.4.2. Binary classification of the frost data . . . 189
13.4.3. Training and test set . . . 189
13.5. Evaluation . . . 189
13.6. ARIMA model selection . . . 190
13.7. Conclusions . . . 192

13.8. Acknowledgments . . . 193
13.9. References . . . 193

**Chapter 14. Efficiency Evaluation of Multiple-Choice Questions and Exams** . . . 195
Evgeny GERSHIKOV and Samuel KOSOLAPOV

14.1. Introduction . . . 195
14.2. Exam efficiency evaluation . . . 196
14.2.1. Efficiency measures and efficiency weighted grades . . . 196
14.2.2. Iterative execution . . . 198
14.2.3. Postprocessing . . . 199
14.3. Real-life experiments and results . . . 200
14.4. Conclusions . . . 203
14.5. References . . . 204

**Chapter 15. Methods of Modeling and Estimation in Mortality** . . . 205
Christos H. SKIADAS and Konstantinos N. ZAFEIRIS

15.1. Introduction . . . 205
15.2. The appearance of life tables . . . 206
15.3. On the law of mortality . . . 207
15.4. Mortality and health . . . 211
15.5. An advanced health state function form . . . 217
15.6. Epilogue . . . 220
15.7. References . . . 221

**Chapter 16. An Application of Data Mining Methods to the Analysis of Bank Customer Profitability and Buying Behavior** . . . 225
Pedro GODINHO, Joana DIAS and Pedro TORRES

16.1. Introduction . . . 225
16.2. Data set . . . 227
16.3. Short-term forecasting of customer profitability . . . 230
16.4. Churn prediction . . . 235
16.5. Next-product-to-buy . . . 236
16.6. Conclusions and future research . . . 238
16.7. References . . . 239

**List of Authors** . . . 241

**Index** . . . 245