

Statistical and Machine-Learning Data Mining

**Techniques for Better Predictive Modeling
and Analysis of Big Data**

Second Edition

Bruce Ratner



CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Contents

| | |
|---|-----------|
| Preface..... | xix |
| Acknowledgments | xxiii |
| About the Author | xxv |
| 1 Introduction..... | 1 |
| 1.1 The Personal Computer and Statistics | 1 |
| 1.2 Statistics and Data Analysis | 3 |
| 1.3 EDA | 5 |
| 1.4 The EDA Paradigm | 6 |
| 1.5 EDA Weaknesses..... | 7 |
| 1.6 Small and Big Data..... | 8 |
| 1.6.1 Data Size Characteristics | 9 |
| 1.6.2 Data Size: Personal Observation of One..... | 10 |
| 1.7 Data Mining Paradigm..... | 10 |
| 1.8 Statistics and Machine Learning | 12 |
| 1.9 Statistical Data Mining..... | 13 |
| References | 14 |
| 2 Two Basic Data Mining Methods for Variable Assessment | 17 |
| 2.1 Introduction | 17 |
| 2.2 Correlation Coefficient | 17 |
| 2.3 Scatterplots | 19 |
| 2.4 Data Mining..... | 21 |
| 2.4.1 Example 2.1 | 21 |
| 2.4.2 Example 2.2 | 21 |
| 2.5 Smoothed Scatterplot..... | 23 |
| 2.6 General Association Test..... | 26 |
| 2.7 Summary | 28 |
| References | 29 |
| 3 CHAID-Based Data Mining for Paired-Variable Assessment | 31 |
| 3.1 Introduction | 31 |
| 3.2 The Scatterplot..... | 31 |
| 3.2.1 An Exemplar Scatterplot..... | 32 |
| 3.3 The Smooth Scatterplot | 32 |
| 3.4 Primer on CHAID | 33 |
| 3.5 CHAID-Based Data Mining for a Smoother Scatterplot | 35 |
| 3.5.1 The Smoother Scatterplot | 37 |

| | | |
|----------|---|-----------|
| 3.6 | Summary | 39 |
| | References | 39 |
| | Appendix | 40 |
| 4 | The Importance of Straight Data: Simplicity and Desirability for Good Model-Building Practice..... | 45 |
| 4.1 | Introduction | 45 |
| 4.2 | Straightness and Symmetry in Data | 45 |
| 4.3 | Data Mining Is a High Concept | 46 |
| 4.4 | The Correlation Coefficient | 47 |
| 4.5 | Scatterplot of (xx3, yy3) | 48 |
| 4.6 | Data Mining the Relationship of (xx3, yy3)..... | 50 |
| | 4.6.1 Side-by-Side Scatterplot | 51 |
| 4.7 | What Is the GP-Based Data Mining Doing to the Data? | 52 |
| 4.8 | Straightening a Handful of Variables and a Baker's Dozen of Variables | 53 |
| 4.9 | Summary | 54 |
| | References | 54 |
| 5 | Symmetrizing Ranked Data: A Statistical Data Mining Method for Improving the Predictive Power of Data..... | 55 |
| 5.1 | Introduction | 55 |
| 5.2 | Scales of Measurement | 55 |
| 5.3 | Stem-and-Leaf Display | 58 |
| 5.4 | Box-and-Whiskers Plot | 58 |
| 5.5 | Illustration of the Symmetrizing Ranked Data Method | 59 |
| | 5.5.1 Illustration 1..... | 59 |
| | 5.5.1.1 Discussion of Illustration 1 | 60 |
| | 5.5.2 Illustration 2..... | 61 |
| | 5.5.2.1 Titanic Dataset | 63 |
| | 5.5.2.2 Looking at the Recoded Titanic Ordinal Variables CLASS_, AGE_, CLASS_AGE_, and CLASS_GENDER_ | 63 |
| | 5.5.2.3 Looking at the Symmetrized-Ranked Titanic Ordinal Variables rCLASS_, rAGE_, rCLASS_AGE_, and rCLASS_GENDER_ | 64 |
| | 5.5.2.4 Building a Preliminary Titanic Model | 66 |
| 5.6 | Summary | 70 |
| | References | 70 |
| 6 | Principal Component Analysis: A Statistical Data Mining Method for Many-Variable Assessment | 73 |
| 6.1 | Introduction | 73 |
| 6.2 | EDA Reexpression Paradigm | 74 |
| 6.3 | What Is the Big Deal?..... | 74 |

| | | |
|----------|--|-----------|
| 6.4 | PCA Basics | 75 |
| 6.5 | Exemplary Detailed Illustration | 75 |
| 6.5.1 | Discussion | 75 |
| 6.6 | Algebraic Properties of PCA | 77 |
| 6.7 | Uncommon Illustration | 78 |
| 6.7.1 | PCA of R_CD Elements ($X_1, X_2, X_3, X_4, X_5, X_6$) | 79 |
| 6.7.2 | Discussion of the PCA of R_CD Elements | 79 |
| 6.8 | PCA in the Construction of Quasi-Interaction Variables | 81 |
| 6.8.1 | SAS Program for the PCA of the Quasi-Interaction Variable | 82 |
| 6.9 | Summary | 88 |
| 7 | The Correlation Coefficient: Its Values Range between Plus/Minus 1, or Do They? | 89 |
| 7.1 | Introduction | 89 |
| 7.2 | Basics of the Correlation Coefficient | 89 |
| 7.3 | Calculation of the Correlation Coefficient | 91 |
| 7.4 | Rematching | 92 |
| 7.5 | Calculation of the Adjusted Correlation Coefficient | 95 |
| 7.6 | Implication of Rematching | 95 |
| 7.7 | Summary | 96 |
| 8 | Logistic Regression: The Workhorse of Response Modeling | 97 |
| 8.1 | Introduction | 97 |
| 8.2 | Logistic Regression Model | 98 |
| 8.2.1 | Illustration | 99 |
| 8.2.2 | Scoring an LRM | 100 |
| 8.3 | Case Study | 101 |
| 8.3.1 | Candidate Predictor and Dependent Variables | 102 |
| 8.4 | Logits and Logit Plots | 103 |
| 8.4.1 | Logits for Case Study | 104 |
| 8.5 | The Importance of Straight Data | 105 |
| 8.6 | Reexpressing for Straight Data | 105 |
| 8.6.1 | Ladder of Powers | 106 |
| 8.6.2 | Bulging Rule | 107 |
| 8.6.3 | Measuring Straight Data | 108 |
| 8.7 | Straight Data for Case Study | 108 |
| 8.7.1 | Reexpressing FD2_OPEN | 110 |
| 8.7.2 | Reexpressing INVESTMENT | 110 |
| 8.8 | Techniques When Bulging Rule Does Not Apply | 112 |
| 8.8.1 | Fitted Logit Plot | 112 |
| 8.8.2 | Smooth Predicted-versus-Actual Plot | 113 |
| 8.9 | Reexpressing MOS_OPEN | 114 |
| 8.9.1 | Plot of Smooth Predicted versus Actual for MOS_OPEN | 115 |

| | | |
|----------|--|-----|
| 8.10 | Assessing the Importance of Variables..... | 118 |
| 8.10.1 | Computing the G Statistic | 119 |
| 8.10.2 | Importance of a Single Variable..... | 119 |
| 8.10.3 | Importance of a Subset of Variables..... | 120 |
| 8.10.4 | Comparing the Importance of Different Subsets of Variables..... | 120 |
| 8.11 | Important Variables for Case Study | 121 |
| 8.11.1 | Importance of the Predictor Variables..... | 122 |
| 8.12 | Relative Importance of the Variables | 122 |
| 8.12.1 | Selecting the Best Subset..... | 123 |
| 8.13 | Best Subset of Variables for Case Study..... | 124 |
| 8.14 | Visual Indicators of Goodness of Model Predictions | 126 |
| 8.14.1 | Plot of Smooth Residual by Score Groups..... | 126 |
| 8.14.1.1 | Plot of the Smooth Residual by Score Groups for Case Study..... | 127 |
| 8.14.2 | Plot of Smooth Actual versus Predicted by Decile Groups | 128 |
| 8.14.2.1 | Plot of Smooth Actual versus Predicted by Decile Groups for Case Study | 129 |
| 8.14.3 | Plot of Smooth Actual versus Predicted by Score Groups | 130 |
| 8.14.3.1 | Plot of Smooth Actual versus Predicted by Score Groups for Case Study | 132 |
| 8.15 | Evaluating the Data Mining Work | 134 |
| 8.15.1 | Comparison of Plots of Smooth Residual by Score Groups: EDA versus Non-EDA Models..... | 135 |
| 8.15.2 | Comparison of the Plots of Smooth Actual versus Predicted by Decile Groups: EDA versus Non-EDA Models | 137 |
| 8.15.3 | Comparison of Plots of Smooth Actual versus Predicted by Score Groups: EDA versus Non-EDA Models | 137 |
| 8.15.4 | Summary of the Data Mining Work | 137 |
| 8.16 | Smoothing a Categorical Variable | 140 |
| 8.16.1 | Smoothing FD_TYPE with CHAID..... | 141 |
| 8.16.2 | Importance of CH_FTY_1 and CH_FTY_2..... | 143 |
| 8.17 | Additional Data Mining Work for Case Study | 144 |
| 8.17.1 | Comparison of Plots of Smooth Residual by Score Group: 4var- versus 3var-EDA Models..... | 145 |
| 8.17.2 | Comparison of the Plots of Smooth Actual versus Predicted by Decile Groups: 4var- versus 3var-EDA Models | 147 |
| 8.17.3 | Comparison of Plots of Smooth Actual versus Predicted by Score Groups: 4var- versus 3var-EDA Models | 147 |

| | | |
|-----------|--|------------|
| 8.17.4 | Final Summary of the Additional Data Mining Work | 150 |
| 8.18 | Summary | 150 |
| 9 | Ordinary Regression: The Workhorse of Profit Modeling | 153 |
| 9.1 | Introduction | 153 |
| 9.2 | Ordinary Regression Model | 153 |
| 9.2.1 | Illustration | 154 |
| 9.2.2 | Scoring an OLS Profit Model | 155 |
| 9.3 | Mini Case Study | 155 |
| 9.3.1 | Straight Data for Mini Case Study | 157 |
| 9.3.1.1 | Reexpressing INCOME | 159 |
| 9.3.1.2 | Reexpressing AGE | 161 |
| 9.3.2 | Plot of Smooth Predicted versus Actual | 162 |
| 9.3.3 | Assessing the Importance of Variables | 163 |
| 9.3.3.1 | Defining the F Statistic and R-Squared | 164 |
| 9.3.3.2 | Importance of a Single Variable | 165 |
| 9.3.3.3 | Importance of a Subset of Variables | 166 |
| 9.3.3.4 | Comparing the Importance of Different Subsets of Variables | 166 |
| 9.4 | Important Variables for Mini Case Study | 166 |
| 9.4.1 | Relative Importance of the Variables | 167 |
| 9.4.2 | Selecting the Best Subset | 168 |
| 9.5 | Best Subset of Variables for Case Study | 168 |
| 9.5.1 | PROFIT Model with gINCOME and AGE | 170 |
| 9.5.2 | Best PROFIT Model | 172 |
| 9.6 | Suppressor Variable AGE | 172 |
| 9.7 | Summary | 174 |
| | References | 176 |
| 10 | Variable Selection Methods in Regression: Ignorable Problem, Notable Solution | 177 |
| 10.1 | Introduction | 177 |
| 10.2 | Background | 177 |
| 10.3 | Frequently Used Variable Selection Methods | 180 |
| 10.4 | Weakness in the Stepwise | 182 |
| 10.5 | Enhanced Variable Selection Method | 183 |
| 10.6 | Exploratory Data Analysis | 186 |
| 10.7 | Summary | 191 |
| | References | 191 |
| 11 | CHAID for Interpreting a Logistic Regression Model | 195 |
| 11.1 | Introduction | 195 |
| 11.2 | Logistic Regression Model | 195 |

| | | |
|-----------|---|------------|
| 11.3 | Database Marketing Response Model Case Study | 196 |
| 11.3.1 | Odds Ratio | 196 |
| 11.4 | CHAID..... | 198 |
| 11.4.1 | Proposed CHAID-Based Method..... | 198 |
| 11.5 | Multivariable CHAID Trees | 201 |
| 11.6 | CHAID Market Segmentation..... | 204 |
| 11.7 | CHAID Tree Graphs | 207 |
| 11.8 | Summary..... | 211 |
| 12 | The Importance of the Regression Coefficient..... | 213 |
| 12.1 | Introduction | 213 |
| 12.2 | The Ordinary Regression Model | 213 |
| 12.3 | Four Questions | 214 |
| 12.4 | Important Predictor Variables..... | 215 |
| 12.5 | P Values and Big Data..... | 216 |
| 12.6 | Returning to Question 1..... | 217 |
| 12.7 | Effect of Predictor Variable on Prediction..... | 217 |
| 12.8 | The Caveat..... | 218 |
| 12.9 | Returning to Question 2..... | 220 |
| 12.10 | Ranking Predictor Variables by Effect on Prediction..... | 220 |
| 12.11 | Returning to Question 3..... | 223 |
| 12.12 | Returning to Question 4..... | 223 |
| 12.13 | Summary..... | 223 |
| | References | 224 |
| 13 | The Average Correlation: A Statistical Data Mining Measure for Assessment of Competing Predictive Models and the Importance of the Predictor Variables | 225 |
| 13.1 | Introduction | 225 |
| 13.2 | Background..... | 225 |
| 13.3 | Illustration of the <i>Difference</i> between Reliability and Validity..... | 227 |
| 13.4 | Illustration of the <i>Relationship</i> between Reliability and Validity..... | 227 |
| 13.5 | The Average Correlation..... | 229 |
| 13.5.1 | Illustration of the Average Correlation with an LTV5 Model | 229 |
| 13.5.2 | Continuing with the Illustration of the Average Correlation with an LTV5 Model..... | 233 |
| 13.5.3 | Continuing with the Illustration with a Competing LTV5 Model | 233 |
| 13.5.3.1 | The Importance of the Predictor Variables..... | 235 |
| 13.6 | Summary..... | 235 |
| | Reference..... | 235 |

| | |
|--|-----|
| 14 CHAID for Specifying a Model with Interaction Variables | 237 |
| 14.1 Introduction | 237 |
| 14.2 Interaction Variables | 237 |
| 14.3 Strategy for Modeling with Interaction Variables | 238 |
| 14.4 Strategy Based on the Notion of a Special Point | 239 |
| 14.5 Example of a Response Model with an Interaction Variable | 239 |
| 14.6 CHAID for Uncovering Relationships | 241 |
| 14.7 Illustration of CHAID for Specifying a Model | 242 |
| 14.8 An Exploratory Look | 246 |
| 14.9 Database Implication | 247 |
| 14.10 Summary | 248 |
| References | 249 |
| | |
| 15 Market Segmentation Classification Modeling with Logistic Regression | 251 |
| 15.1 Introduction | 251 |
| 15.2 Binary Logistic Regression | 251 |
| 15.2.1 Necessary Notation | 252 |
| 15.3 Polychotomous Logistic Regression Model | 253 |
| 15.4 Model Building with PLR | 254 |
| 15.5 Market Segmentation Classification Model | 255 |
| 15.5.1 Survey of Cellular Phone Users | 255 |
| 15.5.2 CHAID Analysis | 256 |
| 15.5.3 CHAID Tree Graphs | 260 |
| 15.5.4 Market Segmentation Classification Model | 263 |
| 15.6 Summary | 265 |
| | |
| 16 CHAID as a Method for Filling in Missing Values | 267 |
| 16.1 Introduction | 267 |
| 16.2 Introduction to the Problem of Missing Data | 267 |
| 16.3 Missing Data Assumption | 270 |
| 16.4 CHAID Imputation | 271 |
| 16.5 Illustration | 272 |
| 16.5.1 CHAID Mean-Value Imputation for a Continuous Variable | 273 |
| 16.5.2 Many Mean-Value CHAID Imputations for a Continuous Variable | 274 |
| 16.5.3 Regression Tree Imputation for LIFE_DOL | 276 |
| 16.6 CHAID Most Likely Category Imputation for a Categorical Variable | 278 |
| 16.6.1 CHAID Most Likely Category Imputation for GENDER | 278 |
| 16.6.2 Classification Tree Imputation for GENDER | 280 |
| 16.7 Summary | 283 |
| References | 284 |

| | |
|--|-----|
| 17 Identifying Your Best Customers: Descriptive, Predictive, and Look-Alike Profiling | 285 |
| 17.1 Introduction | 285 |
| 17.2 Some Definitions..... | 285 |
| 17.3 Illustration of a Flawed Targeting Effort..... | 286 |
| 17.4 Well-Defined Targeting Effort..... | 287 |
| 17.5 Predictive Profiles | 290 |
| 17.6 Continuous Trees | 294 |
| 17.7 Look-Alike Profiling | 297 |
| 17.8 Look-Alike Tree Characteristics..... | 299 |
| 17.9 Summary..... | 301 |
| 18 Assessment of Marketing Models | 303 |
| 18.1 Introduction | 303 |
| 18.2 Accuracy for Response Model..... | 303 |
| 18.3 Accuracy for Profit Model..... | 304 |
| 18.4 Decile Analysis and Cum Lift for Response Model..... | 307 |
| 18.5 Decile Analysis and Cum Lift for Profit Model..... | 308 |
| 18.6 Precision for Response Model..... | 310 |
| 18.7 Precision for Profit Model | 312 |
| 18.7.1 Construction of SWMAD | 314 |
| 18.8 Separability for Response and Profit Models | 314 |
| 18.9 Guidelines for Using Cum Lift, HL/SWMAD, and CV..... | 315 |
| 18.10 Summary..... | 316 |
| 19 Bootstrapping in Marketing: A New Approach for Validating Models | 317 |
| 19.1 Introduction | 317 |
| 19.2 Traditional Model Validation | 317 |
| 19.3 Illustration..... | 318 |
| 19.4 Three Questions | 319 |
| 19.5 The Bootstrap..... | 320 |
| 19.5.1 Traditional Construction of Confidence Intervals | 321 |
| 19.6 How to Bootstrap | 322 |
| 19.6.1 Simple Illustration | 323 |
| 19.7 Bootstrap Decile Analysis Validation | 325 |
| 19.8 Another Question | 325 |
| 19.9 Bootstrap Assessment of Model Implementation Performance | 327 |
| 19.9.1 Illustration..... | 330 |
| 19.10 Bootstrap Assessment of Model Efficiency | 331 |
| 19.11 Summary..... | 334 |
| References | 336 |

| | | |
|-----------|---|-----|
| 20 | Validating the Logistic Regression Model: Try Bootstrapping | 337 |
| 20.1 | Introduction | 337 |
| 20.2 | Logistic Regression Model..... | 337 |
| 20.3 | The Bootstrap Validation Method | 337 |
| 20.4 | Summary | 338 |
| | Reference | 338 |
| 21 | Visualization of Marketing ModelsData Mining to Uncover Innards of a Model | 339 |
| 21.1 | Introduction | 339 |
| 21.2 | Brief History of the Graph | 339 |
| 21.3 | Star Graph Basics..... | 341 |
| 21.3.1 | Illustration..... | 342 |
| 21.4 | Star Graphs for Single Variables | 343 |
| 21.5 | Star Graphs for Many Variables Considered Jointly..... | 344 |
| 21.6 | Profile Curves Method | 346 |
| 21.6.1 | Profile Curves Basics..... | 346 |
| 21.6.2 | Profile Analysis | 347 |
| 21.7 | Illustration..... | 348 |
| 21.7.1 | Profile Curves for RESPONSE Model | 350 |
| 21.7.2 | Decile Group Profile Curves | 351 |
| 21.8 | Summary | 354 |
| | References | 355 |
| | Appendix 1: SAS Code for Star Graphs for Each Demographic Variable about the Deciles..... | 356 |
| | Appendix 2: SAS Code for Star Graphs for Each Decile about the Demographic Variables | 358 |
| | Appendix 3: SAS Code for Profile Curves: All Deciles | 362 |
| 22 | The Predictive Contribution Coefficient: A Measure of Predictive Importance | 365 |
| 22.1 | Introduction | 365 |
| 22.2 | Background | 365 |
| 22.3 | Illustration of Decision Rule..... | 367 |
| 22.4 | Predictive Contribution Coefficient..... | 369 |
| 22.5 | Calculation of Predictive Contribution Coefficient..... | 370 |
| 22.6 | Extra Illustration of Predictive Contribution Coefficient..... | 372 |
| 22.7 | Summary | 376 |
| | Reference | 377 |
| 23 | Regression Modeling Involves Art, Science, and Poetry, Too | 379 |
| 23.1 | Introduction | 379 |
| 23.2 | Shakespearean Modelogue..... | 379 |

| | | |
|-----------|--|------------|
| 23.3 | Interpretation of the Shakespearean Modelogue | 380 |
| 23.4 | Summary | 384 |
| | References | 384 |
| 24 | Genetic and Statistic Regression Models: A Comparison | 387 |
| 24.1 | Introduction | 387 |
| 24.2 | Background | 387 |
| 24.3 | Objective | 388 |
| 24.4 | The GenIQ Model, the Genetic Logistic Regression | 389 |
| 24.4.1 | Illustration of "Filling up the Upper Deciles" | 389 |
| 24.5 | A Pithy Summary of the Development of Genetic Programming | 392 |
| 24.6 | The GenIQ Model: A Brief Review of Its Objective and Salient Features | 393 |
| 24.6.1 | The GenIQ Model Requires Selection of Variables and Function: An Extra Burden? | 393 |
| 24.7 | The GenIQ Model: How It Works | 394 |
| 24.7.1 | The GenIQ Model Maximizes the Decile Table | 396 |
| 24.8 | Summary | 398 |
| | References | 398 |
| 25 | Data Reuse: A Powerful Data Mining Effect of the GenIQ Model | 399 |
| 25.1 | Introduction | 399 |
| 25.2 | Data Reuse | 399 |
| 25.3 | Illustration of Data Reuse | 400 |
| 25.3.1 | The GenIQ Profit Model | 400 |
| 25.3.2 | Data-Reused Variables | 402 |
| 25.3.3 | Data-Reused Variables GenIQvar_1 and GenIQvar_2 | 403 |
| 25.4 | Modified Data Reuse: A GenIQ-Enhanced Regression Model | 404 |
| 25.4.1 | Illustration of a GenIQ-Enhanced LRM | 404 |
| 25.5 | Summary | 407 |
| 26 | A Data Mining Method for Moderating Outliers Instead of Discarding Them | 409 |
| 26.1 | Introduction | 409 |
| 26.2 | Background | 409 |
| 26.3 | Moderating Outliers Instead of Discarding Them | 410 |
| 26.3.1 | Illustration of Moderating Outliers Instead of Discarding Them | 410 |
| 26.3.2 | The GenIQ Model for Moderating the Outlier | 414 |
| 26.4 | Summary | 414 |

| | | |
|-----------|---|-----|
| 27 | Overfitting: Old Problem, New Solution | 415 |
| 27.1 | Introduction..... | 415 |
| 27.2 | Background..... | 415 |
| | 27.2.1 Idiomatic Definition of Overfitting to Help Remember the Concept..... | 416 |
| 27.3 | The GenIQ Model Solution to Overfitting..... | 417 |
| | 27.3.1 RANDOM_SPLIT GenIQ Model..... | 420 |
| | 27.3.2 RANDOM_SPLIT GenIQ Model Decile Analysis..... | 420 |
| | 27.3.3 Quasi N-tile Analysis..... | 422 |
| 27.4 | Summary..... | 424 |
| 28 | The Importance of Straight Data: Revisited | 425 |
| 28.1 | Introduction..... | 425 |
| 28.2 | Restatement of Why It Is Important to Straighten Data..... | 425 |
| 28.3 | Restatement of Section 9.3.1.1 “Reexpressing INCOME”..... | 426 |
| | 28.3.1 Complete Exposition of Reexpressing INCOME..... | 426 |
| | 28.3.1.1 The GenIQ Model Detail of the gINCOME Structure..... | 427 |
| 28.4 | Restatement of Section 4.6 “Data Mining the Relationship of (xx3, yy3)”..... | 428 |
| | 28.4.1 The GenIQ Model Detail of the GenIQvar(yy3) Structure..... | 428 |
| 28.5 | Summary..... | 429 |
| 29 | The GenIQ Model: Its Definition and an Application | 431 |
| 29.1 | Introduction..... | 431 |
| 29.2 | What Is Optimization?..... | 431 |
| 29.3 | What Is Genetic Modeling?..... | 432 |
| 29.4 | Genetic Modeling: An Illustration..... | 434 |
| | 29.4.1 Reproduction..... | 437 |
| | 29.4.2 Crossover..... | 437 |
| | 29.4.3 Mutation..... | 438 |
| 29.5 | Parameters for Controlling a Genetic Model Run..... | 440 |
| 29.6 | Genetic Modeling: Strengths and Limitations..... | 441 |
| 29.7 | Goals of Marketing Modeling..... | 442 |
| 29.8 | The GenIQ Response Model..... | 442 |
| 29.9 | The GenIQ Profit Model..... | 443 |
| 29.10 | Case Study: Response Model..... | 444 |
| 29.11 | Case Study: Profit Model..... | 447 |
| 29.12 | Summary..... | 450 |
| | Reference..... | 450 |
| 30 | Finding the Best Variables for Marketing Models | 451 |
| 30.1 | Introduction..... | 451 |
| 30.2 | Background..... | 451 |

| | | |
|-----------|--|------------|
| 30.3 | Weakness in the Variable Selection Methods | 453 |
| 30.4 | Goals of Modeling in Marketing | 455 |
| 30.5 | Variable Selection with GenIQ | 456 |
| 30.5.1 | GenIQ Modeling | 459 |
| 30.5.2 | GenIQ Structure Identification | 460 |
| 30.5.3 | GenIQ Variable Selection | 463 |
| 30.6 | Nonlinear Alternative to Logistic Regression Model | 466 |
| 30.7 | Summary | 469 |
| | References | 470 |
| 31 | Interpretation of Coefficient-Free Models | 471 |
| 31.1 | Introduction | 471 |
| 31.2 | The Linear Regression Coefficient | 471 |
| 31.2.1 | Illustration for the Simple Ordinary Regression Model | 472 |
| 31.2.2 | Illustration for the Simple Logistic Regression Model | 473 |
| 31.3 | The Quasi-Regression Coefficient for Simple Regression Models | 474 |
| 31.3.1 | Illustration of Quasi-RC for the Simple Ordinary Regression Model | 474 |
| 31.3.2 | Illustration of Quasi-RC for the Simple Logistic Regression Model | 475 |
| 31.3.3 | Illustration of Quasi-RC for Nonlinear Predictions | 476 |
| 31.4 | Partial Quasi-RC for the Everymodel | 478 |
| 31.4.1 | Calculating the Partial Quasi-RC for the Everymodel | 480 |
| 31.4.2 | Illustration for the Multiple Logistic Regression Model | 481 |
| 31.5 | Quasi-RC for a Coefficient-Free Model | 487 |
| 31.5.1 | Illustration of Quasi-RC for a Coefficient-Free Model | 488 |
| 31.6 | Summary | 494 |
| | Index | 497 |