## Contents

Preface xv

1 Misinterpretation of Selected Theoretical Concepts of Statistics 1
  1.1 Introduction 1
  1.2 What is a Population? 2
  1.3 A Sample and Sample Space 2
    1.3.1 What is a Sample? 2
    1.3.2 What is the Sample Space? 3
    1.3.3 What is a Representative Sample? 6
    1.3.4 Relationship between the Sample Space, Population, and a Sample 7
  1.4 Distribution of a Random Sample Space 8
  1.5 What is a Random Variable? 9
  1.6 Theoretical Concept of a Random Sample 9
    1.6.1 What is a Random Sample in Statistics? 9
    1.6.2 Central Limit Theorem 10
    1.6.3 Unbiased Statistics based on Random Samples 16
    1.6.4 Special Notes on Nonrandom Sample 19
  1.7 Does a Representative Sample Really Exist? 19
  1.8 Remarks on Statistical Powers and Sample Sizes 21
  1.9 Hypothesis and Hypothesis Testing 24
  1.10 Groups of Research Variables 25
    1.10.1 Problem Indicators 26
    1.10.2 Controllable Cause Factors 26
    1.10.3 Uncontrollable Cause Factors 26
    1.10.4 Background or Classification Factors 27
    1.10.5 Environmental Factors 27
  1.11 Causal Relationship between Variables 27
    1.11.1 Bivariate Correlation 27
    1.11.2 Special Remarks 30
  1.12 Misinterpretation of Selected Statistics 31
    1.12.1 Standard Error 31
    1.12.2 Significance Level and Power of a Test 31
    1.12.3 Reliability of a Test or Instrument 32
2 Simple Statistical Analysis but Good for Strategic Decision Making 37
2.1 Introduction 37
2.2 A Single Input for Decision Making 39
   2.2.1 A Single Sampled Unit 39
   2.2.2 Descriptive Statistics Based on a Single Measurable Variable 39
   2.2.3 Agung Six-Point Scale (ASPS) Problem Indicator 43
   2.2.4 Latent Variables and Composite Indexes 45
   2.2.5 Demographic and Social–Economic Factors 45
   2.2.6 Garbage as a Data Source 46
   2.2.7 Boxplot as an Input for Decision Making 46
   2.2.8 A Series of Inputs for Strategic Decision Making 48
2.3 Data Transformation 48
   2.3.1 To Generate Categorical Variables 49
   2.3.2 To Generate Dummy Variables 51
2.4 Biserial Correlation Analysis 51
2.5 One-Way Tabulation of a Variable 53
2.6 Two-Way Tabulations 54
   2.6.1 Measure of Associations for Bivariate Categorical Variables 58
   2.6.2 Other Measures of Association Based on a $2 \times 2$ Table 58
   2.6.3 Measures of Association Based on a $1 \times 2$ Table 64
2.7 Three-Way Tabulation 67
   2.7.1 Conditional Measures of Association for a $2 \times 2 \times 2$ Table 69
   2.7.2 Conditional Odds Ratio for an $I \times J \times 2$ Table 70
2.8 Special Notes and Comments 74
2.9 Special Cases of the N-Way Incomplete Tables 77
2.10 Partial Associations 80
2.11 Multiple Causal Associations Based on Categorical Variables 81
   2.11.1 Theoretical and Empirical Concepts of Causal Associations 81
   2.11.2 Multidimensional Frequency Table 85
2.12 Seemingly Causal Model Based on Categorical Variables 89
   2.12.1 Causal Association Based on $(X_1, X_2, Y_1)$ or $(X_1, Y_1, Y_2)$ 90
   2.12.2 Causal Association Based on $(X_1, X_2, Y_1, Y_2)$ 91
   2.12.3 Causal Association Based on Multidimensional Variables 94
2.13 Alternative Descriptive Statistical Summaries 95
   2.13.1 Application of the Object “Descriptive Statistics and Test” 95
   2.13.2 Application of the Object “Graph…” 102
2.14 How to Present Descriptive Statistical Summary? 107
   2.14.1 DSS Based on a Set of Zero–One Indicators 107
   2.14.2 Two-Dimensional DSS of Proportions 108
   2.14.3 Multidimensional DSS of Proportions 108
   2.14.4 DSS Based on a Set of Agung–Likert Scale Attributes 109
   2.14.5 DSS Based on a Set of Numerical Problem Indicators 110
   2.14.6 Additional Descriptive Statistical Summaries 111
2.15 General Seemingly Causal Model 111
2.16 Empirical Studies Presenting Descriptive Statistical Summaries 112
  2.16.1 Studies in the Field of Nutrition 112
  2.16.2 Studies in Public Health 114
  2.16.3 Selected Experimental Studies 114
  2.16.4 Studies in Public Relations 114
  2.16.5 Studies on Other Population Problems 115

3 One-Way Proportion Models 117
  3.1 Introduction 117
  3.2 One-Way Proportion Models Based on a $2 \times 2$ Table 117
    3.2.1 Regression Functions 118
    3.2.2 Binary Logit Functions 119
    3.2.3 Odds Ratio Statistics 120
  3.3 Binary Choice Models Based on a $K \times 2$ Table 121
    3.3.1 Binary Logit Models 121
    3.3.2 Binary Multiple Regressions 122
  3.4 Binary Logit Models Based on $N$-Way Tabulation 122
    3.4.1 Binary Logit Models Based on Three-Way Tabulation 122
    3.4.2 Binary Choice Models Based on Higher Dimensional Tables 124
  3.5 General Binary Choice Models 124
    3.5.1 Binary Multiple Regression Model 125
    3.5.2 The Wald Test 127
    3.5.3 Binary Logit Models 134
    3.5.4 Binary Probit Models 144
    3.5.5 Binary Extreme-Value Models 147
  3.6 Special Notes and Comments 151
    3.6.1 The True Population Binary Choice Model 151
    3.6.2 The Sampled Binary Choice Function 151
    3.6.3 Alternative Equation Estimations 152
  3.7 Association between Categorical Variables 152
    3.7.1 Generating the Dummy Variables 153
    3.7.2 Generating a Cell Factor 154
  3.8 One-Way Binary Choice Models Based on $N$-Way Tabulation 156
    3.8.1 $N$-Way Tabulation without an Empty Cell 156
    3.8.2 $N$-Way Tabulation with Empty Cells 157
    3.8.3 Testing Hypotheses 157
  3.9 Special Notes and Comments on Binary Choice Models 160

4 $N$-Way Cell-Proportion Models 165
  4.1 Introduction 165
  4.2 The $N$-Way Tabulation of Proportions 165
    4.2.1 A $2 \times 2$ Table of Proportions 165
    4.2.2 A $I \times J$ Table of Proportions 167
  4.3 The $2 \times 2$ Factorial Model of Proportions 168
    4.3.1 Pure Interaction Models 168
    4.3.2 Interaction Models with a Main Factor 170
### 4.3.3 Interaction Models with Both Main Factors 174
### 4.3.4 Additive Binary Choice Models 175
### 4.4 $I \times J$ Factorial Models of Proportions 176
- 4.4.1 Interaction Models 176
- 4.4.2 Special Notes and Comments 178
### 4.5 Multifactorial Cell-Proportion Model 180
### 4.6 Presenting the Statistical Summary 188

## 5 N-Way Cell-Mean Models 193
### 5.1 Introduction 193
### 5.2 One-Way Multivariate Cell-Mean Models 195
- 5.2.1 An MCMM without an Intercept 195
- 5.2.2 An MCMM with Intercepts 195
### 5.3 N-Way Multivariate Cell-Mean Models 197
- 5.3.1 Two-Way Multivariate Cell-Mean Models 197
- 5.3.2 Three-Way Multivariate Cell-Mean Model 201
- 5.3.3 N-Way Multivariate Cell-Mean Model 202
### 5.4 Equality Test by Classification 202
### 5.5 Testing Weighted Means Differences 208
### 5.6 Descriptive Statistical Summary 212

## 6 Multinomial Choice Models with Categorical Exogenous Variables 213
### 6.1 Introduction 213
### 6.2 Multinomial Choice Models 213
- 6.2.1 Multinomial Logit Model as a Set of $(M-1)$ Binary Logit Models 213
- 6.2.2 Multinomial Logit Model as a Set of M Binary Choice Models 224
### 6.3 Ordered Choice Models 225
- 6.3.1 Simple Ordered Choice Models 225
### 6.4 Concordance–Discordance Measure of Association 231
### 6.5 Multifactorial Ordered Choice Models 234
### 6.6 Multilevel Choice Models 241
- 6.6.1 Two-Level Choice Models 241
- 6.6.2 Three-Level Choice Models 250
### 6.7 Special Notes on the Multinomial Logit Model 253
### 6.8 Selected Population Studies Using Multinomial Choice Models 256
- 6.8.1 Multinomial Problem Indicators and Gender Equity Indexes 256
- 6.8.2 Multinomial Problem and Poverty Indicators 259

## 7 General Choice Models 263
### 7.1 Introduction 263
### 7.2 Binary Choice Models with a Numerical Variable 263
- 7.2.1 The Simplest Binary Choice Model 263
- 7.2.2 Alternative Simple Binary Choice Models 269
- 7.2.3 Special Notes and Comments 276
### 7.3 Heterogeneous Binary Choice Models 276
- 7.3.1 The Simplest Heterogeneous Binary Choice Model 276
- 7.3.2 General Heterogeneous Binary Choice Model 282
7.4 Homogeneous Binary Choice Models 284
  7.4.1 Binary Choice ANCOVA Model with a Numerical Variable 284
  7.4.2 Graphical Representation of an ANCOVA Model 287
7.5 General Binary Choice Models 288
  7.5.1 Hierarchical Binary Logit Model 288
  7.5.2 Nonhierarchical Binary Logit Model 289
  7.5.3 Additive Binary Logit Model 290
  7.5.4 GBCM with Two Numerical and a Dichotomous Independent Variable 293
  7.5.5 GBCM with Two Numerical and a Set of Categorical Independent Variables 297
7.6 Advanced Binary Choice Models 298
  7.6.1 Binary Choice Heterogeneous Regressions 298
  7.6.2 Binary Choice ANCOVA Model 304
  7.6.3 Descriptive Statistical Summaries 307
7.7 Multidimensional Binary Choice Translog Linear Model 307
7.8 Piecewise Binary Choice Models 309
7.9 Ordered Choice Models with Numerical Independent Variables 313
7.10 Studies Using General Choice Models 325
7.11 Two-Stage Binary Choice Model 326

8 Experimental Data Analysis 329
  8.1 Introduction 329
  8.2 Analysis Based on Cell-Mean Models 329
    8.2.1 The Simplest Statistical Analysis 330
    8.2.2 Special Remarks 331
    8.2.3 Application of Multivariate Cell-Mean Models 332
  8.3 Bivariate Correlation Analysis 333
  8.4 Effects of the Experimental Factors 334
  8.5 Effects of the Experimental Factors and Covariates 335
    8.5.1 Effects of the Experimental Factors and a Covariate 336
    8.5.2 Effects of the Experimental Factors and Two Covariates 342
    8.5.3 The Application of Translog Linear Models 346
  8.6 Application of the Ordered Choice Models 356
  8.7 Application of Seemingly Causal Models 360
    8.7.1 The Simplest Seemingly Causal Model 361
    8.7.2 Four Pairs of Causal Relationships 363
    8.7.3 Five Pairs of Causal Relationships 364
    8.7.4 All Pairs Have Causal Relationships 365
    8.7.5 Alternative Seemingly Causal Models 368
    8.7.6 Special Notes and Comments 369
  8.8 Multivariate Analysis of Covariance 369
  8.9 Tests for Equality of Medians 372
  8.10 The Simplest Experimental Design 376
9 Seemingly Causal Models Based on Numerical Variables 381
9.1 Introduction 381
9.2 The Simplest Seemingly Causal Model 382
  9.2.1 Bivariate Correlation and the Simplest Linear Regression 382
  9.2.2 Scatter Graph with Regression Line 385
  9.2.3 Residual Analysis 389
  9.2.4 Special Notes and Comments 390
9.3 General Linear Models Based on Bivariate (X, Y) 391
  9.3.1 Continuous Regression Models 391
  9.3.2 Discontinuous Regressions 402
  9.3.3 Regressions by a Classification Factor 405
9.4 Models Based on Numerical Trivariate 407
  9.4.1 Continuous Regressions 407
  9.4.2 Regressions by Classification Factors 416
9.5 Regression Analysis Using the Principal Components 417
9.6 Seemingly Causal Models Based on (X1, X2, Y1, Y2) 420
9.7 Seemingly Causal Models Based on (X1, X2, X3, Y1, Y2) 422
  9.7.1 The Model with the Dependent Variable Y1 423
  9.7.2 The Model with the Dependent Variable Y2 424
  9.7.3 The Model with the Dependent Variable X1 424
  9.7.4 The Model with the Dependent Variable X3 424
9.8 New Types of Interaction Model 426
  9.8.1 Polynomial Interaction Model 426
  9.8.2 General Polynomial Interaction Model 428
  9.8.3 System Polynomial Interaction Model 430
9.9 Special Cases 431
  9.9.1 Predicted Variables and Predictors 431
  9.9.2 The Simplest and the Most Complex Seemingly
       Causal Models 433
9.10 Special Notes and Comments 434
Appendix A.9.1 Hypothetical Data Set 435

10 Factor Analysis and Latent Variables Models 439
10.1 Introduction 439
10.2 The Basic Concept of Factor Analysis 440
10.3 The First-Level Latent Variables 441
  10.3.1 Generating Additive Latent Variables 441
  10.3.2 Interaction Latent Variables 447
  10.3.3 Special Notes and Comments 449
10.4 Illustrations Based on Hamssal’s (2006) Data Set 450
  10.4.1 Generating Latent Variables 450
  10.4.2 Latent Variable Regression Models 452
  10.4.3 Alternative Latent Variable Models 453
10.5 Selected Cases Based on Ary Suta’s (2006) Data Set 458
  10.5.1 Multilevel Latent Variables 458
  10.5.2 Problems with the Sample Sizes 459
10.6 Evaluation Analysis Based on Latent Variables 462
  10.6.1 Ordinal Classification Based on a Latent Variable 462
  10.6.2 Composite Index Based on a Latent Variable 463
  10.6.3 N-Way Tabulation Based on Latent Variables 464

11 Application of the Stepwise Selection Methods 467
  11.1 Introduction 467
  11.2 The Options for the Stepwise Selection Methods 467
  11.3 Selection Method for the Numerical Variable Regression Models 469
    11.3.1 Two-Way Interaction Stepwise Regressions 469
    11.3.2 Three-Way Interaction Stepwise Regressions 473
    11.3.3 Application of Multistage Stepwise Selections 476
    11.3.4 Alternative Selection Methods 478
  11.4 Multifactorial Stepwise Regression Models 480
    11.4.1 Multifactorial Cell-Mean Models 480
    11.4.2 Multifactorial Heterogeneous Regressions 482
    11.4.3 Stepwise ANCOVA Models 491
  11.5 Illustrative Stepwise Regressions Based on Mlogit.wfl 495
    11.5.1 Classical ANCOVA Models 495
    11.5.2 Interaction ANCOVA Models 497
  11.6 Special Notes and Comments 503

12 Censored Multiple Regression Models 505
  12.1 Introduction 505
  12.2 Tobit Models 505
    12.2.1 Tobit Cell-Mean Models 506
    12.2.2 Tobit Regression Models with Numerical
        Independent Variables 511
    12.2.3 Selected Studies Using Tobit Regressions 517
  12.3 General Tobit Model 517
  12.4 Zero–One Indicator of Censoring 521
  12.5 Illustrative Cases of Censored Observations 526
    12.5.1 Outliers are Considered as Censored Observations 526
    12.5.2 Both Tales of Observations are Considered as
        Censored Observations 527
    12.5.3 Waiting Time and Switching Status Variables 528
  12.6 Series for a Censoring Variable 528
  12.7 Switching Censored Regressions 531
  12.8 Special Notes and Comments 542
  Appendix A.12.1 Hypothetical Censored Data, Modified from
  Faad’s (2008) Data Set 543

References 545

Index 551