



From *Applying Data Science*.
Full book available for purchase [here](#).

Contents

Dedication	iii
About This Book	ix
About The Author	xxvix
Acknowledgments	xxxix
Case Study 1 – Performing Headcount Survival Analysis for Employee Retention	1
Chapter 1: Using Survival Analysis Methods to Analyze Employee Retention Time	3
1.1 Introduction	3
1.1.1 Time-to-Event Data	3
1.1.2 Analytical Methods for Time-to-Event Data	4
1.2 Overview of the Case Study.....	4
1.3 Business Background and Business Question.....	4
1.3.1 Business Background.....	4
1.3.2 Business Questions.....	5
1.3.3 Employee Retention Data.....	5
1.4 Simple Descriptive Statistics Do Not help	7
1.5 The Kaplan-Meier Method Can Deal with Censored Data	8
1.5.1 The Basic Idea	8
1.5.2 Analyzing the Individual Duration	9
1.5.3 Code Example.....	10
1.5.4 Graphical Representation of the Kaplan-Meier Curve	11
1.6 Detailed Analysis of the Survival Curve.....	12
1.6.1 Creating the Survival Curve for All Employees	12
1.6.2 Interpreting the Survival Curve	14
1.6.3 Adding Confidence Bands to the Survival Curve	15
1.7 Interpreting the Hazard Curve.....	16
1.7.1 Basic Idea of the Hazard Curve	16
1.7.2 Adding a Plot for the Hazard Curve.....	16
1.7.3 The Hazard Curve for the SALES_ENGINEER Department	17
1.8 Additional Methods in PROC LIFETEST	18
1.8.1 Using the Lifetable Method	18
1.8.2 Generating an Output Data Set.....	20
1.9 Conclusion.....	21
Chapter 2: Analyzing the Effect of Influential Factors on Employee Retention Time	23

2.1 Introduction	23
2.2 Analyzing the Employee Data by Department	23
2.2.1 Descriptive Results	23
2.2.2 Survival Analysis.....	25
2.3 Additional Stratified Analyses.....	31
2.3.1 Survival Analysis by Gender and Technical Knowledge	31
2.3.2 The Misleading Effect of Left Truncated Data	33
2.4 Quantifying the Effect of Influential Variables	34
2.4.1 The Cox Proportional Hazards Regression	34
2.4.2 Results of the Cox Proportional Hazards Regression.....	36
2.4.3 Explained Variation of the Cox Proportional Hazards Model	37
2.4.4 Creating Output Data Sets	38
2.5 Preparing Time-to-Event Data.....	40
2.5.1 General Points	40
2.5.2 Business Decisions for the Definition of Events.....	40
2.6 Other Procedures in SAS/STAT® for the Analysis of Time-to-Event Data.....	41
2.7 Conclusion.....	42
Chapter 3: Performing Survival Data Mining - The Data Mining Approach for Survival Analysis.....	43
3.1 Introduction	43
3.2 The Idea of Survival Data Mining.....	44
3.2.1 Using a Multinomial Logistic Regression	44
3.2.2 Expanding the Data	44
3.3 Using SAS Enterprise Miner for Survival Data Mining	45
3.3.1 Using the Survival Node	45
3.3.2 Survival Data Mining Example - Preparation.....	45
3.3.3 Survival Data Mining Example – Results.....	48
3.4 Expanding Data for Survival Data Mining.....	52
3.4.1 Rationale	52
3.4.2 Different Data Requirements between SAS Enterprise Miner and SAS/STAT Procedures.....	52
3.4.3 Two Types of “Expanded” Data.....	52
3.5 Code Example to Expand Data for Survival Data Mining	54
3.5.1 General	54
3.5.2 Using a DATA Step Loop to Expand the EMPLOYEES Data.....	54
3.5.3 Merging a Multiple-Row-per-Subject Data Set to Expand the Employees Data	55
3.5.4 Be Careful with the Alignment of the Data!	57
3.6 Conclusion.....	58
Chapter 4: Visualizing Employee Retention Data	59
4.1 Introduction and Overview.....	59
4.2 Initial Preparations for the Reports.....	60
4.2.1 General Points	60
4.2.2 Initialization Statements	60

4.3 The Career-Start-End Plot	60
4.3.1 General Idea	60
4.3.2 The Graph and Its Interpretation	61
4.3.3 Data Preparation for the Career-Start-End Plot	61
4.4 Employees-Win-Loss-Plot	63
4.4.1 General Idea	63
4.4.2 The Graph and Its Interpretation	63
4.4.3 Preparing the data	64
4.5 Cumulated-Knowledge Plot	67
4.5.1 General Idea	67
4.5.2 The Graph and Its Interpretation	67
4.5.3 Preparing the Data	69
4.6 Conclusion	70
Case Study 2 – Detecting Structural Changes and Outliers in Longitudinal Data	71
Chapter 5: Analyzing and Smoothing the Course of Longitudinal Data	73
5.1 Introduction	73
5.2 Analyzing Longitudinal Data	74
5.2.2 Methods for Detection of Changes in Longitudinal Data	74
5.2.3 Automatic Detection of Changes and Outliers	75
5.3 Airline Passengers Data from 1990 and 2004	75
5.3.1 The Data	75
5.3.2 Business Questions	76
5.4 Considerations When Smoothing Longitudinal Data	77
5.4.1 Information Reduction	77
5.4.2 Length of the Smoothing Window	77
5.4.3 Weighting of the Periods	77
5.4.4 Orientation of the Smoothing Window	77
5.5 Smoothing Your Data with the EXPAND Procedure	78
5.5.1 General	78
5.5.2 Backward Smoothing	78
5.5.3 Centered Smoothing	79
5.5.4 Specifying the Moving Weights Manually	81
5.5.5 Trimming in Case of Incomplete Data	81
5.6 Smoothing Longitudinal Data with a SAS DATA Step	84
5.6.1 Rationale for a SAS DATA Step Implementation	84
5.6.2 Brute-Force Method	84
5.6.3 Efficient Method	85
5.6.4 Calculating a Centered Moving Average with a DATA Step	86
5.6.5 Using a User-Defined SAS Function	86
5.7 Conclusion	87
Chapter 6: Detecting Structural Changes in Longitudinal Data	89
6.1 Introduction	89
6.2 Detect Structural Changes with Analytical Methods	90
6.2.1 Multivariate Adaptive Regression Splines	90

6.2.2 Automatic Detection of Knot Values	90
6.3 Using the ADAPTIVEREG Procedure	92
6.3.1 General	92
6.3.2 Basic Usage Examples	92
6.4 Breakpoints in the Airline Passenger Data	94
6.4.1 Interpreting the Breakpoints	94
6.4.2 Displaying the Breakpoints in the Line Chart.....	96
6.5 Structural Changes in Clinical Trial Data	98
6.5.1 Business Background and Data	98
6.5.2 Patient Recruitment over Time	99
6.5.3 Baseline Characteristics in a Clinical Trial	101
6.6 Conclusion.....	102
Chapter 7 – Detecting Outliers and Level Shifts in Longitudinal Data	105
7.1 Introduction	105
7.2 Detecting Outliers and Level Shifts	106
7.2.1 Overview.....	106
7.2.2 Level Shifts.....	108
7.2.3 Outliers	108
7.2.4 Other Event Types.....	109
7.3 Using SAS for the Detection of Events in Longitudinal Data	109
7.3.1 Procedures in SAS/ETS® Software.....	109
7.3.2 Methods in SAS® Forecast Server	110
7.4 Event Detection in the Airline Passenger Data.....	111
7.4.1 Overview.....	111
7.4.2 Detecting Events with the X13 Procedure.....	111
7.4.3 Creating a Plot with Reference Lines.....	112
7.5 Other Methods for Event Detection	115
7.5.1 Overview.....	115
7.5.2 Detecting Events with the X11 Procedure.....	115
7.5.2 Detecting Events with the HPFDIAGNOSE Procedure.....	116
7.6 Conclusion.....	117
Chapter 8 – Results from a Simulation Study with Longitudinal Data	119
8.1 Introduction	119
8.2 Simulating Longitudinal Data	119
8.2.1 Properties of the Simulated Data	119
8.2.2 Preparations for Simulating the Data.....	120
8.2.3 Using a DATA Step to Generate the Data.....	121
8.3 Detecting Structural Changes in the Simulated Data	123
8.3.1 Analyzing the Original Data	123
8.3.2 Detecting Changes in the Smoothed Data	124
8.4 Detecting Events in the Simulated Data.....	126
8.4.1 Using the X12 Procedure.....	126
8.4.2 Interpretation	127

8.5 Conclusion.....	128
Chapter 9 – Analyzing the Variability of Longitudinal Data.....	129
9.1 Introduction	129
9.2 Analyzing Variability over Time.....	130
9.2.1 Visual Interpretation of the Data.....	130
9.2.2 The Effect of Too Much Variability	131
9.3 Monitoring Longitudinal Data with Control Charts	132
9.3.1 Method from Statistical Process Control	132
9.3.2 Detecting Structural Changes in the X- and S- Curves.....	133
9.4 Misleading Changes and Outliers	135
9.4.1 Alternating Differences between March and April.....	135
9.4.2 Decline in Sales Numbers between January and April	136
9.5 Conclusion.....	136
9.6 Conclusion of This Case Study	137
Case Study 3 – Explaining Forecast Errors and Deviations	139
Chapter 10 – Investigating Forecast Errors with Descriptive Statistics	141
10.1 Introduction	141
10.2 Business Questions and Background	142
10.2.1 The Business Environment of the Case Study	142
10.2.2 Measuring the Forecast Error	144
10.2.3 Business Questions for the Analysis.....	144
10.3 Data Preparation.....	145
10.3.1 Available Source Data.....	145
10.3.2 Merging the Tables Together.....	147
10.3.3 Calculating Derived Variables.....	148
10.5 Descriptive Analysis of the Forecast Errors	150
10.5.1 Overview.....	150
10.5.2 Checking the Distribution of the Forecast Error.....	150
10.5.3 Using Box Plots to Analyze the Forecast Error by Subgroup.....	153
10.5.4 Analyzing Target Year and Model Type	158
10.5.5 Band Charts with Quartiles	161
10.6 Conclusion.....	163
Chapter 11 – Investigating Forecast Errors with General Linear Models.....	165
11.1 Introduction	165
11.2 Regression Analysis in SAS	165
11.2.1 Procedures in SAS/STAT®	165
11.2.2 General and Generalized Might Confuse You	166
11.2.3 Similar Procedures and Their Focus	166
11.3 Regression Analysis	167
11.3.1 General Points on Regression Models.....	167
11.3.2 Univariate Regression Analysis	169
11.4 Multivariate Regression Analysis	172
11.4.1 Using the GLMSELECT Procedure	172

11.4.2 Results from the Regression Model.....	173
11.4.3 Comparison with the Log-Transformed Model.....	174
11.4.4 Coefficients of the Variables in the Regression Model.....	174
11.5 Scoring New Observations with the GLMSELECT Procedure.....	176
11.6 Conclusion.....	178
Chapter 12 – Interpreting the Coefficients of Categorical Variables in Regression Models	179
12.1 Introduction.....	179
12.2 Categorical Variables and Dummy Variables.....	179
12.2.1 Problem Description.....	179
12.2.2 Different Types of Dummy Variable Coding.....	180
12.2.3 Defining Dummy Variables in Selected SAS Regression Procedures.....	181
12.3 Displaying Regression Coefficients for All Categories.....	183
12.3.1 Introduction.....	183
12.3.2 Usage Example.....	184
12.3.3 Macro Code Explained Step-by-Step.....	185
12.3.4 Summary.....	189
12.4 Imputing Missing Values with the GLM Procedure.....	189
12.4.1 Missing Values.....	189
12.4.2 A Code Example for Imputing Missing Values.....	190
12.5 Further Topics for SAS Regression Procedure.....	192
12.5.1 Saving Tables with the ODS OUTPUT Statement.....	192
12.5.2 Changing the Selection Statistic for the Stepwise Selection.....	193
12.5.3 Creating a Multi-Step Model.....	193
12.5 Conclusion.....	194
Chapter 13: Using Quantile Regression to Get More Than the Average Picture ...	195
13.1 Introduction.....	195
13.2 Basic Idea of Quantile Regression.....	196
13.2.1 Introduction.....	196
13.2.2 SAS Procedures for Quantile Regression.....	197
13.2.3 Robustness of the Quantile Regression.....	198
13.2.4 Model for the 0.25 Quantile Compared.....	200
13.3 Quantile Regression for the Statistical Forecast Error.....	202
13.3.1 Introduction.....	202
13.3.2 Quantile Regression for Selected Quartiles.....	202
13.3.3 Using the HPQUANTSELECT Procedure.....	204
13.3.4 Creating a Process Plot for the Parameter Estimates.....	205
13.4 Sampling of Data.....	206
13.4.1 Overview.....	206
13.4.2 Sampling with the SAS DATA Step.....	207
13.4.3 Sampling with the SURVEYSELECT Procedure.....	207
13.5 Conclusion.....	208
Chapter 14 – Analyzing the Effect of Manual Overrides in Forecasting.....	209

14.1 Introduction	209
14.2 Manual Overrides.....	209
14.2.1 Available Data.....	209
14.2.2 Quantifying the Effect of Manual Overrides	210
14.3 Univariate Results.....	212
14.3.1 Differences between Judgmental and Statistical Forecast.....	212
14.3.2 Analyzing Selected Variables.....	216
14.4 Results from the Regression Model	219
14.4.1 Introduction.....	219
14.4.2 Investigating Only the Size and Direction of the Manual Overwrite	219
14.4.3 Adding More Explanatory Factors.....	221
14.5 Conclusion.....	224
Case Study 4 – Forecasting the Demand for New Products	225
Chapter 15 – Performing Demand Forecasting for New Products	227
15.1 Introduction to the Case Study	227
15.2 Introduction	228
15.3 Business Questions and Background	228
15.3.1 Definition of Short Demand History	228
15.3.2 Reasons for Short or No Demand History	229
15.3.3 New Product or Short-Term Product.....	229
15.4 Analytic Approaches for Forecasting New Products.....	230
15.4.1 Overview.....	230
15.4.2 Demand Forecasting Using Predictive Modeling	230
15.4.3 Using Similarity Search for New Product Forecasting.....	230
15.4.4 Deriving Forecasts from Similarity Clusters.....	230
15.4.5 Using Time Series Similarity Analysis	230
15.5 Data Preparation Steps	231
15.5.1 Available Data	231
15.5.2 Generated Output Data.....	232
15.5.3 Overview of the Data Preparation Steps	232
15.5.4 Number of Products Launched in the Same Month	233
15.5.5 Aggregating the Demand Data	235
15.6 Creating the Demand Data Mart	236
15.6.1 Overview.....	236
15.6.2 Basic Variables	236
15.6.3 Adding the Known Demand	237
15.7 Adding Known Demand Data to Product Base Data	239
15.7.1 Introduction.....	239
15.7.2 Demand Data in One Row per Product.....	239
15.8 Conclusion.....	240
Chapter 16 – Using Poisson Regression to Forecast the Demand for New Products.....	241
16.1 Introduction.....	241

16.2 Poisson Regression.....	241
16.2.1 Generalized Linear Model.....	241
16.2.2 Implementation in SAS	242
Overview.....	242
16.3 Demand Forecasting Using Predictive Modeling	242
16.3.1 Basic Idea.....	242
16.3.2 Analysis Data	242
16.3.3 Independency of Observations.....	243
16.4 Poisson Regression Using the HPGENSELECT Procedure	244
16.4.1 SAS Code	244
16.4.2 Results of the Poisson Regression	245
16.4.3 Extending the Model Syntax	247
16.5 Scoring a New Product	248
16.5.1 Data for the New Product.....	248
16.5.2 Applying the Score Logic	249
16.5.3 Forecast Report.....	249
16.5.4 Graphical Output	250
16.6 Conclusion.....	251
Chapter 17 – Using Similarity Search to Forecast the Demand for New Products.....	253
17.1 Introduction	253
17.2 Basic Idea of Similarity Search.....	253
17.2.1 Basic Idea.....	253
17.2.2 Advantage of Similarity Search.....	254
17.3 Implementing Similarity Search	254
17.3.1 Product Repository	254
17.3.2 New Product Data	254
17.3.3 Defining the Weights.....	255
17.3.4 Similarity Search with the SQL Procedure	255
17.3.5 Preparation of the Results.....	257
17.3.6 Displaying the Demand Data.....	258
17.4 Extending the Similarity Search	260
17.4.1 Creating a SAS Stored Process	260
17.4.2 Similarity Search for Existing Products	261
17.5 Conclusion.....	261
Case Study 5 – Checking the Alignment with Predefined Pattern.....	263
Chapter 18 – Checking Accounting Data for the Benford’s Law.....	265
18.1 Introduction to the Case Study	265
18.1.1 General Idea and Examples.....	265
18.1.2 Statistical Implementation.....	266
18.2 Business Questions and Background	266

18.2.1 The Business Environment of the Case Study	266
18.2.2 The Available Data and Their Background	267
18.2.3 Business Questions for the Analysis of Benford's Law	268
18.3 Benford's Law	269
18.3.1 General Idea	269
18.3.2 Calculating the Expected Benford Proportions	269
18.3.3 Discussion of the Benford's Law	270
18.4 Checking Data for the Benford's Law	271
18.4.1 Using the Chi ² -Distribution to Check Against the Expected Distribution	271
18.4.2 Deriving the First Digit from the Amount Value	272
18.4.3 Using the FREQ Procedure	272
18.5 Results	274
18.5.1 Preparation	274
18.5.2 Graphical Output	275
18.5.3 Tabular Output	275
18.5.4 Interpretation	276
18.6 Consideration of Missing Digits	276
18.6.1 Problem Description	276
18.6.2 Accounting for Zero Frequencies	277
18.6.3 Example for Data with No Digit 8	278
18.7 Conclusion	279
Chapter 19 – Checking the Benford's Law for Multiple Accounts	281
19.1 Introduction	281
19.2 Running Benford Analysis in BY-Entity Mode	281
19.2.1 Basic Idea	281
19.2.2 Implementation with the FREQ Procedure	282
19.2.3 Preparing a Table with Chi ² -Values	283
19.3 Graphical Display of the Results	284
19.3.1 Overall Distribution of p-Values	284
19.3.2 Displaying Individual Courses	285
19.4 Accounting for Zero Frequencies	287
19.4.1 Problem Description	287
19.4.2 Using the TIMESERIES Procedure	287
19.5 Conclusion	289
Chapter 20 – Checking Different Patterns in the Data	291
20.1 Introduction	291
20.2 General Idea	291
20.2.1 Overview	291
20.2.2 Bank Account Data	292
20.3 Analysis by Time Period	292
20.3.1 Overall Results	292
20.3.2 Analysis per Year	293
20.3.3 Analysis per Month	294
20.3.4 Interpretation	296

20.4 Matching a Predefined Distribution	297
20.4.1 General Idea	297
20.4.2 Measures for Sales Targets	298
20.4.3 Ranking the Sales Agents	299
20.4.4 Investigating Individual Account Managers	301
20.5 Course of Customer Usage Data over Time	303
20.5.1 General Idea	303
20.5.2 Example Output	303
20.5.3 Summary	304
20.6 Conclusion	304

Case Study 6 – Listening to Your Data – Discover Relationships with Unsupervised Analysis Methods 305

Chapter 21 – Finding Relationships in Your Analysis Data with Association Analysis 307

21.1 General Idea	307
21.1.3 Methods in SAS for Unsupervised Data Analysis	308
21.2 Business Question and Example Data	309
21.2.1 Association Analysis	309
21.2.2 Insurance Claim Data.....	309
21.3 Creating a Feature Data Mart from One-Row-Per-Subject Analysis Data	311
21.3.1 Overview	311
21.3.2 Using a SAS DATA Step.....	311
21.3.3 Alternate Approach: Using the TRANSPOSE Procedure	312
21.4 Performing Association Analysis with SAS® Enterprise Miner	314
21.4.1 Creating the SAS® Enterprise Miner Process Flow	314
21.4.2 Viewing the Results.....	316
21.4.3 Adding the p-Value for the Associations	318
21.5 Interpreting the Results.....	319
21.5.1 General	319
21.5.2 Interpreting the Rules	320
21.5.3 Obvious Results.....	320
21.5.4 Distant Associations	321
21.5.5 Analyzing Specific Subgroups	321
21.5.6 Visual Analysis of the Data	322
21.6 Conclusion.....	323

Chapter 22 – Using Variable Clustering to Detect Relationships in Your Data 325

22.1 Introduction	325
22.2 Variable Clustering	326
22.2.1 Idea of Variable Clustering	326
22.2.2 Example Data.....	326
22.3 Preparing the Data.....	327
22.3.1 Create a Data Set with Indicator Variables	327

22.3.2 Method 1: Using the TRANSPOSE Procedure.....	327
22.3.3 Method 2: Using a SAS DATA Step.....	329
22.3.4 Method 3: Using the LOGISTIC Procedure.....	329
22.3.5 Conclusion	330
22.4 Performing Variable Clustering with the VARCLUS Procedure.....	330
22.4.1 General	330
22.4.2 Using the VARCLUS Procedure	331
22.4.3 Results of the VARCLUS Procedure	331
22.4.4 Creating a Tree Diagram	332
22.5 Interpretation of the Results.....	334
22.5.1 Reading the Tree Diagram	334
22.5.2 Creating and Reading the Variables Table	335
22.5.3 Interactive Analysis	335
22.6 Conclusion.....	336
Chapter 23 – Investigating of Clinical Trial Data in an Explanatory Way	337
23.1 Introduction	337
23.2 Patient Data from a Long-Term Clinical Trial	337
23.2.1 Basic Idea.....	337
23.2.2 Feature Categories in the Data.....	338
23.3 Preparing the Data for Association Analysis	339
23.3.1 Overview.....	339
23.3.2 Creating the Categories	339
23.3.3 Creating a Table with Features per Patient.....	341
23.4 Results of the Association Analysis.....	341
23.4.1 Obvious Rules Because of the Definition of the Data	342
23.4.2 Intuitive Rules from a Medical Point of View.....	342
23.4.3 Analyzing Specific Subgroups	343
23.5 Performing Variable Clustering	345
23.5.1 Introduction.....	345
23.5.2 Preparing the Data for Variable Clustering	345
23.5.3 Performing Variable Clustering.....	346
23.5.4 Interpretation of the Results	347
23.6 Conclusion.....	348
Case Study 7 – Using Monte Carlo Simulations to Understand the Outcome Distribution	349
Chapter 24 – Calculating the Outcomes of All Possible Scenarios.....	351
24.1 Overview over the Case Study	351
24.1.1 Introduction.....	351
24.1.2 Sales Manager’s Pipeline	352
24.2 Example Data and Business Question	352
24.2.1 The List of Sales Projects.....	352
24.2.2 Expectations of Upper Management	353
24.3 Some Basic Calculations	354

24.3.1 Not to Be Applied! – Incorrect Quick Calculation	354
24.3.2 Calculating the Expected Value	354
24.4 Complete Calculation	355
24.4.1 The Basic Idea	355
24.4.2 Some Basic Considerations	355
24.4.3 Looking at More Projects	356
24.5 Performing the Complete Calculation with SAS/IML Software	357
24.5.1 Introduction.....	357
24.5.2 General Idea of Using SAS IML for the Complete Calculations	357
24.5.3 Full SAS/IML Code	358
24.5.4 Details for the SAS/IML Code	358
24.6 Results from the Complete Calculation	362
24.6.1 Introduction.....	362
24.6.2 Results.....	363
24.6.3 Interpretation of the Results	364
24.7 Using a SAS DATA Step for the Complete Calculations.....	365
24.7.1 General Idea.....	365
24.7.2 SAS Code	365
24.8 Conclusion.....	368
Chapter 25 – Using Monte Carlo Methods to Simulate the Distribution of the Outcomes	369
25.1 Introduction	369
25.2 Performing Monte Carlo Simulations	369
25.2.1 Alternative to the Complete Calculation.....	369
25.2.2 Basic Idea.....	370
25.3 Using SAS/IML Software to Perform the Monte Carlo Simulations	370
25.3.1 Input Data.....	370
25.3.2 Full SAS/IML Code	370
25.3.3 SAS/IML Code in Details.....	371
25.4 Results of the Monte Carlo Simulations.....	373
25.4.1 Creating the Analysis Reports	373
25.4.2 Results of Different Scenarios	374
25.5 Using a SAS DATA Step to Perform the Monte Carlo Simulations.....	376
25.5.1 General Idea.....	376
25.5.2 The SAS DATA Step Code	377
25.6 Conclusion.....	378
Case Study 8 – Studying Complex Systems – Simulating the Monopoly Board Game	379
Chapter 26 – Creating a Basic Framework to Simulate the Visit Frequency on the Fields of the Monopoly Board Game	381
26.1 Simulating Complex Systems.....	381
26.1.1 Understanding Complex Systems with Simulation Studies	381

26.1.2 Case Study: Monopoly Board Game	382
26.2 Overview of the Case Study.....	382
26.3 Description of the Monopoly Board Game.....	382
26.3.1 Introduction.....	382
26.3.2 The Idea of the Game.....	382
26.3.3 Assumptions for the Simulation Runs.....	384
26.4 Business Questions of Interest	385
26.5 Simulating the Monopoly Board Game with a SAS DATA Step	386
26.5.1 Overview.....	386
26.5.2 Simulation Architecture	386
26.6 The SAS DATA Step for the Simulation in Detail	387
26.6.1 Initialization	387
26.6.2 Looping over Scenarios (Games).....	389
26.6.3 Looping over Rounds and Players.....	389
26.6.4 Considering the Go-to-Jail Directive	390
26.6.5 Closing the Macro Loop	391
26.7 Results from the Simulation Runs	391
26.7.1 Results in the Data Set	391
26.7.2 Preparing the Analysis Data.....	392
26.7.3 Graphical Output of the Players' Positions	393
26.7.4 Tabular Output of the Players' Positions	394
26.7.5 Considering the "Go-to-Jail" Directive	395
26.8 Conclusion.....	396
Chapter 27 – Enhancing the Simulation Framework to Consider Special Rules	397
27.1 Overview	397
27.2 Considering the Chance and Community Chest Fields.....	397
27.2.1 Business Background.....	397
27.2.2 Considering Community Chest Fields	398
27.2.3 Considering Chance Fields	398
27.2.4 SAS Code to Generate the Results	399
27.2.5 Results and Interpretation.....	401
27.3 Considering the Speed Die	402
27.3.1 Instructions for the Speed Die	402
27.3.2 Effect of the Relocation Rule	403
27.3.3 Implementation of the Speed Die Rules	403
27.4 Results of Considering the Speed Die.....	405
27.4.1 Distribution of the Visit Frequency	405
27.4.2 Illustrating the Dynamic over the Course of the Game	407
27.5 Conclusion.....	411
Chapter 28 – Simulating the Profitability of the Property Fields of the Monopoly Board Game.....	413
28.1 Introduction	413
28.2 Relation to Business Decisions.....	413
28.3 Tracking Players' Positions and Transactions.....	414

28.3 Implementation with a SAS DATA Step	417
28.3.1 Using a SAS Format as Lookup Table.....	417
28.3.2 Calculating the Monetary Values in the Simulation	418
28.4 Field Profit over the Course of the Game	420
28.4.1 Preparing the Data	420
28.4.2 Results and Interpretation.....	421
28.5 Analyzing the Final Field Profit and Field Setup	422
28.5.1 Introduction.....	422
28.5.2 Field Profit	422
28.5.3 Field Setup	424
28.6 Profit and Risk Analysis	425
28.6.1 General Idea.....	425
28.6.2 Creation of the Results	425
28.6.3 Interpretation of the Results	426
28.7 Conclusion	427
Appendix A – SAS Macros	429
A.1 The %VAREXIST macro	429
A.2 The %DetectBrkPoint Macro	430
A.3 The %CALC_REFERENCE_CATEGORY Macro.....	432
A.4 The %REPLACE_MV Macro	434
A.5 The %Monopoly_Sim Macro	435
References	441
Recommended Reading.....	443
Index	445