



From *Business Survival Analysis Using SAS*.  
Full book available for purchase [here](#).

# Contents

<b>About The Book</b> .....	<b>vii</b>
<b>About the Author</b> .....	<b>xiii</b>
<b>Acknowledgments</b> .....	<b>xv</b>
<b>Contents</b> .....	<b>iii</b>
<b>Chapter 1: Data Preparation for Survival Models</b> .....	<b>1</b>
Introduction .....	1
Step 1: Collect Three Raw Variables .....	3
Step 2: Create Three Required Variables .....	3
Step 3: Collect and Process Many Predictor Variables .....	3
Data Set Structure .....	3
Characteristics of the Data Set Used in This Book .....	4
Special Considerations for Survival Analysis Data Input .....	4
Plan for the Collection of the Required Variables .....	4
Five Fundamental Questions for Time-to-Next-Purchase Survival Models .....	4
Data Set Creation .....	7
(Start, End) Calendar Timeline .....	7
Time and Censor Variables .....	9
Strata by Previous Purchases .....	12
Categorical Variables .....	13
Data Preparation with Base SAS, Macros, and SAS/STAT .....	14
Multiple Imputation .....	14
Winsorization .....	15
Binning Using Macros .....	15
Data Preparation with SAS Enterprise Miner .....	16
<b>Chapter 2: Exploratory Data Analysis and Basic Statistical Concepts</b> .....	<b>19</b>
Introduction .....	19
Overview of Continuous and Discrete Distributions .....	20
Deciding Whether to Use Continuous Time or Discrete Time .....	21
Understanding the Concept of a Customer Tie .....	22
Continuous Distribution .....	24
Survival Function for Continuous Time .....	24
Hazard Function for Continuous Time .....	27
Discrete Distribution .....	28

Data Formats .....	29
Customer-Level Data Set (Standard Data Format) .....	29
Calculation of the Discrete Hazard—An Example .....	29
Interpretation of the Results .....	31
Month Data Set (Expanded Data Format) .....	32
Creation of Dummy Variables .....	34
A Simple Example.....	34
A Time-to-Next-Purchase Example.....	35
PROC LOGISTIC to Compute Hazard.....	38
Odds Calculation (January) .....	39
Hazard Calculation (January) .....	40
Survival Function .....	41
Hazard and Survival Relationship .....	42
Summary of Distributions .....	43
Life Table Method .....	43
Step 1: PROC LIFETEST with METHOD = LIFE .....	44
Step 2: At-Risk Adjustment .....	49
Step 3: Empirical Hazard Calculation.....	49
Step 4: Graphics and Interpretation .....	50
Data Storage Formats .....	52
Customer-Level Data Set (Standard Format) .....	52
Month Data Set (Expanded Format) .....	52
<b>Chapter 3: Stratified Sampling Method for Survival Analysis .....</b>	<b>53</b>
Introduction .....	53
The Sampling Process with PROC SURVEYSELECT .....	54
Step 1: Analyze Sample Proportion for Strata.....	55
Step 2: Create the Model_Shopping Sample.....	56
Step 3: Create the Score_Shopping Sample .....	57
Step 4: Compare the Results from the Model_Shopping and Score_Shopping Data Sets .....	59
SAS Enterprise Miner Nodes .....	60
The Sample Node .....	60
The Data Partition Node .....	60
The Flowchart .....	61
<b>Chapter 4: SAS Enterprise Miner Use from Source to Customer-Level Model Output.....</b>	<b>63</b>
Introduction .....	63
Creating the Process Flow.....	64
Step 1: Create a New Project .....	65
Step 2: Create a New Library .....	66
Step 3: Create Start Code.....	67
Step 4: Create a New Data Source .....	69
Step 5: Create a Process Flow Diagram .....	74

Step 6: Insert the Data Partition Node .....	76
Step 7: Create the Survival Process Flow .....	80
Running a Model and Examining Results .....	81
Survival Node Properties Panel .....	81
Results .....	83
<b>Chapter 5: The Cubic Spline Regression and Model Interpretation .....</b>	<b>85</b>
Introduction .....	85
SAS Enterprise Miner Strategy for Standard Data Format .....	86
The Problem: The Nonlinearity of the Hazard Function .....	87
The Solution: The Cubic Spline Basis Functions .....	87
Two Challenges: Number of Knots to Choose and Placement of the Knots .....	89
The Mathematical Definition of a Knot.....	90
The Default Five-Knots Cubic Spline Basis Method without Stepwise Selection .....	92
The Default Five-Knots Cubic Splines Basis Method Using the Stepwise Selection.....	95
Example 1: Interpretation of Numeric Variables.....	97
Understanding the Stepwise Regression Spline Model in SAS Enterprise Miner .....	98
Writing the Model .....	100
Understanding the Odds Ratio.....	100
Example 2: Interpretation of Categorical Variables.....	105
Odds Calculation of Categorical Variables.....	105
Interpretations .....	107
Model Output.....	111
Model Information .....	111
Strata Analysis .....	116
Survival and Hazard Functions Saved in a Data Set.....	121
<b>Chapter 6: The Fully Expanded Data Set Model .....</b>	<b>123</b>
Introduction .....	123
Saving the Expanded Data Set .....	123
Include a SAS Code Node.....	123
Rename the SAS Code Node.....	124
Create a Permanent Library .....	125
Explore the SURV_EXPCENDATA Data Set .....	126
Running the Expanded Model.....	128
Create a Second Diagram and Assign Variable Roles .....	128
Set the Properties and Run .....	129
Comparing the Results.....	130
<b>Chapter 7: Statistics for Performance and Validation .....</b>	<b>131</b>
Introduction .....	131
Twofold Focus .....	132
SAS Enterprise Miner for Practical Solutions.....	132
Comparison of Model Validation Approaches .....	133
Validation: An Explanation.....	133
Traditional Logistic Regression .....	134
Survival Regression (Steps 1 through 4).....	136

Statistics for Performance and Validation.....	141
Step 5: Compute the Hazard for Only the First Month of the Period .....	141
Step 6: Create a Hit Variable Indicating Whether Repurchase Occurs in January.....	142
Step 7: Compute the Depth of the Hazard in Quantiles .....	143
Step 8: Compute the Sensitivity .....	146
Step 9: Create the Concentration Curve .....	147
Step 10: Compute Lift .....	149
Step 11: Create the Lift Curve .....	150
Step 12: Compute Benefit.....	152
Step 13: Create the Benefit Curve .....	152
Step 14: Select the Cutoff=Depth Where the Benefit Is Maximum .....	153
The Depth Measure and the Hazard Measure: Which to Choose.....	155
Two Business Situations.....	155
Business Situation 1: Use of Depth as Cutoff .....	156
Business Situation 2: Use of Average Hazard Function Value as Cutoff.....	158
Specificity and Suboptimal Cutoffs .....	159
Gini Concentration Ratio .....	163
Kolmogorov-Smirnov.....	165
Method 1—PROC NPAR1WAY .....	167
Method 2—PROC SQL .....	168
Density Analysis and PROC KDE .....	170
A Comprehensive Validation Approach .....	172
<b>Chapter 8: Scoring New and Old Customers.....</b>	<b>173</b>
Introduction .....	173
Business Situation 1: Scoring Customers .....	175
Preparing the Scoring Data Set .....	175
Scoring Using SAS Enterprise Miner.....	178
Analyzing the Optimized Score Code.....	190
Scoring Using SAS DATA Steps.....	197
Analyzing Customer ID = 4 .....	199
Using PROC COMPARE for Validation.....	203
Business Situation 2: Calibrating Scoring Results .....	205
Step 1: Compute Actual and Forecast of Part_TRAIN .....	207
Step 2: Plot an Actual and Forecast Graph—Part_TRAIN .....	208
Step 3: Compute the Calibration (Prior Proportion—Part_TRAIN) .....	209
Step 4: Compute the Full-Year Forecast—Part_VALIDATE .....	209
Step 5: Calibrate the Forecast of Part_VALIDATE .....	211
Step 6: Plot an Actual by Calibrated Forecast Graph .....	212
<b>References .....</b>	<b>213</b>
<b>Index .....</b>	<b>215</b>