

Contents

1	Introduction	1
1.1	What Is Survival Analysis?	1
1.2	What You Need to Know to Use This Book	2
1.3	Survival Data and Censoring	2
1.4	Some Examples of Survival Data Sets	6
1.5	Additional Notes	9
2	Basic Principles of Survival Analysis	11
2.1	The Hazard and Survival Functions	11
2.2	Other Representations of a Survival Distribution	13
2.3	Mean and Median Survival Time	14
2.4	Parametric Survival Distributions	15
2.5	Computing the Survival Function from the Hazard Function	19
2.6	A Brief Introduction to Maximum Likelihood Estimation	20
2.7	Additional Notes	23
3	Nonparametric Survival Curve Estimation	25
3.1	Nonparametric Estimation of the Survival Function	25
3.2	Finding the Median Survival and a Confidence Interval for the Median	30
3.3	Median Follow-Up Time	32
3.4	Obtaining a Smoothed Hazard and Survival Function Estimate	32
3.5	Left Truncation	36
3.6	Additional Notes	41
4	Nonparametric Comparison of Survival Distributions	43
4.1	Comparing Two Groups of Survival Times	43
4.2	Stratified Tests	49
4.3	Additional Note	52

5	Regression Analysis Using the Proportional Hazards Model	55
5.1	Covariates and Nonparametric Survival Models	55
5.2	Comparing Two Survival Distributions Using a Partial Likelihood Function	56
5.3	Partial Likelihood Hypothesis Tests	59
5.3.1	The Wald Test	60
5.3.2	The Score Test	60
5.3.3	The Likelihood Ratio Test	60
5.4	The Partial Likelihood with Multiple Covariates	63
5.5	Estimating the Baseline Survival Function	64
5.6	Handling of Tied Survival Times	65
5.7	Left Truncation	69
5.8	Additional Notes	71
6	Model Selection and Interpretation	73
6.1	Covariate Adjustment	73
6.2	Categorical and Continuous Covariates	74
6.3	Hypothesis Testing for Nested Models	78
6.4	The Akaike Information Criterion for Comparing Non-nested Models	81
6.5	Including Smooth Estimates of Continuous Covariates in a Survival Model	84
6.6	Additional Note	86
7	Model Diagnostics	87
7.1	Assessing Goodness of Fit Using Residuals	87
7.1.1	Martingale and Deviance Residuals	87
7.1.2	Case Deletion Residuals	92
7.2	Checking the Proportion Hazards Assumption	94
7.2.1	Log Cumulative Hazard Plots	94
7.2.2	Schoenfeld Residuals	96
7.3	Additional Note	100
8	Time Dependent Covariates	101
8.1	Introduction	101
8.2	Predictable Time Dependent Variables	106
8.2.1	Using the Time Transfer Function	107
8.2.2	Time Dependent Variables That Increase Linearly with Time	109
8.3	Additional Note	110
9	Multiple Survival Outcomes and Competing Risks	113
9.1	Clustered Survival Times and Frailty Models	113
9.1.1	Marginal Survival Models	115
9.1.2	Frailty Survival Models	116
9.1.3	Accounting for Family-Based Clusters in the "ashkenazi" Data	117

9.1.4	Accounting for Within-Person Pairing of Eye Observations in the Diabetes Data	120
9.2	Cause-Specific Hazards	121
9.2.1	Kaplan-Meier Estimation with Competing Risks	121
9.2.2	Cause-Specific Hazards and Cumulative Incidence Functions	123
9.2.3	Cumulative Incidence Functions for Prostate Cancer Data	126
9.2.4	Regression Methods for Cause-Specific Hazards	127
9.2.5	Comparing the Effects of Covariates on Different Causes of Death	130
9.3	Additional Notes	134
10	Parametric Models	137
10.1	Introduction	137
10.2	The Exponential Distribution	138
10.3	The Weibull Model	138
10.3.1	Assessing the Weibull Distribution as a Model for Survival Data in a Single Sample	138
10.3.2	Maximum Likelihood Estimation of Weibull Parameters for a Single Group of Survival Data	141
10.3.3	Profile Weibull Likelihood	142
10.3.4	Selecting a Weibull Distribution to Model Survival Data	143
10.3.5	Comparing Two Weibull Distributions Using the Accelerated Failure Time and Proportional Hazards Models	146
10.3.6	A Regression Approach to the Weibull Model	148
10.3.7	Using the Weibull Distribution to Model Survival Data with Multiple Covariates	149
10.3.8	Model Selection and Residual Analysis with Weibull Survival Data	151
10.4	Other Parametric Survival Distributions	153
10.5	Additional Note	154
11	Sample Size Determination for Survival Studies	157
11.1	Power and Sample Size for a Single Arm Study	157
11.2	Determining the Probability of Death in a Clinical Trial	161
11.3	Sample Size for Comparing Two Exponential Survival Distributions	163
11.4	Sample Size for Comparing Two Survival Distributions Using the Log-Rank Test	165
11.5	Determining the Probability of Death from a Non-parametric Survival Curve Estimate	166
11.6	Example: Calculating the Required Number of Patients for a Randomized Study of Advanced Gastric Cancer Patients	169

11.7	Example: Calculating the Required Number of Patients for a Randomized Study of Patients with Metastatic Colorectal Cancer	170
11.8	Using Simulations to Estimate Power	171
11.9	Additional Notes	174
12	Additional Topics	177
12.1	Using Piecewise Constant Hazards to Model Survival Data	177
12.2	Interval Censoring	187
12.3	The Lasso Method for Selecting Predictive Biomarkers	192
	Erratum	E1
A	A Basic Guide to Using R for Survival Analysis	201
A.1	The R System	201
A.1.1	A First R Session	202
A.1.2	Scatterplots and Fitting Linear Regression Models	204
A.1.3	Accommodating Non-linear Relationships	207
A.1.4	Data Frames and the Search Path for Variable Names	209
A.1.5	Defining Variables Within a Data Frame	211
A.1.6	Importing and Exporting Data Frames	211
A.2	Working with Dates in R	212
A.2.1	Dates and Leap Years	213
A.2.2	Using the "as.date" Function	213
A.3	Presenting Coefficient Estimates Using Forest Plots	215
A.4	Extracting the Log Partial Likelihood and Coefficient Estimates from a coxph Object	217
	References	218
	Index	223
	R Package Index	225