

Contents

List of Tables	xiii
List of Figures	xv
Preface to the Third Edition	xix
Preface to the Second Edition	xxi
Preface to the Revised Edition	xxiii
Preface to the First Edition	xxv
Notation and Typography	xxvii
1 The problem of survival analysis	1
1.1 Parametric modeling	2
1.2 Semiparametric modeling	3
1.3 Nonparametric analysis	5
1.4 Linking the three approaches	5
2 Describing the distribution of failure times	7
2.1 The survivor and hazard functions	7
2.2 The quantile function	10
2.3 Interpreting the cumulative hazard and hazard rate	13
2.3.1 Interpreting the cumulative hazard	13
2.3.2 Interpreting the hazard rate	15
2.4 Means and medians	16
3 Hazard models	19
3.1 Parametric models	20
3.2 Semiparametric models	21
3.3 Analysis time (time at risk)	24

4	Censoring and truncation	29
4.1	Censoring	29
4.1.1	Right-censoring	30
4.1.2	Interval-censoring	32
4.1.3	Left-censoring	34
4.2	Truncation	34
4.2.1	Left-truncation (delayed entry)	34
4.2.2	Interval-truncation (gaps)	35
4.2.3	Right-truncation	36
5	Recording survival data	37
5.1	The desired format	37
5.2	Other formats	40
5.3	Example: Wide-form snapshot data	44
6	Using stset	47
6.1	A short lesson on dates	48
6.2	Purposes of the stset command	51
6.3	Syntax of the stset command	51
6.3.1	Specifying analysis time	52
6.3.2	Variables defined by stset	55
6.3.3	Specifying what constitutes failure	57
6.3.4	Specifying when subjects exit from the analysis	59
6.3.5	Specifying when subjects enter the analysis	62
6.3.6	Specifying the subject-ID variable	65
6.3.7	Specifying the begin-of-span variable	67
6.3.8	Convenience options	70
7	After stset	73
7.1	Look at stset's output	73
7.2	List some of your data	76
7.3	Use stdescribe	77
7.4	Use stvary	78

7.5	Perhaps use <code>stfill</code>	80
7.6	Example: Hip fracture data	82
8	Nonparametric analysis	91
8.1	Inadequacies of standard univariate methods	91
8.2	The Kaplan–Meier estimator	93
8.2.1	Calculation	93
8.2.2	Censoring	96
8.2.3	Left-truncation (delayed entry)	97
8.2.4	Interval-truncation (gaps)	99
8.2.5	Relationship to the empirical distribution function	99
8.2.6	Other uses of <code>sts list</code>	101
8.2.7	Graphing the Kaplan–Meier estimate	102
8.3	The Nelson–Aalen estimator	107
8.4	Estimating the hazard function	113
8.5	Estimating mean and median survival times	117
8.6	Tests of hypothesis	122
8.6.1	The log-rank test	123
8.6.2	The Wilcoxon test	125
8.6.3	Other tests	125
8.6.4	Stratified tests	126
9	The Cox proportional hazards model	129
9.1	Using <code>stcox</code>	130
9.1.1	The Cox model has no intercept	131
9.1.2	Interpreting coefficients	131
9.1.3	The effect of units on coefficients	133
9.1.4	Estimating the baseline cumulative hazard and survivor functions	135
9.1.5	Estimating the baseline hazard function	139
9.1.6	The effect of units on the baseline functions	143

9.2	Likelihood calculations	145
9.2.1	No tied failures	145
9.2.2	Tied failures	148
	The marginal calculation	148
	The partial calculation	149
	The Breslow approximation	150
	The Efron approximation	151
9.2.3	Summary	151
9.3	Stratified analysis	152
9.3.1	Obtaining coefficient estimates	152
9.3.2	Obtaining estimates of baseline functions	155
9.4	Cox models with shared frailty	156
9.4.1	Parameter estimation	157
9.4.2	Obtaining estimates of baseline functions	161
9.5	Cox models with survey data	164
9.5.1	Declaring survey characteristics	165
9.5.2	Fitting a Cox model with survey data	166
9.5.3	Some caveats of analyzing survival data from complex survey designs	168
9.6	Cox model with missing data—multiple imputation	169
9.6.1	Imputing missing values	171
9.6.2	Multiple-imputation inference	173
10	Model building using <code>stcox</code>	177
10.1	Indicator variables	177
10.2	Categorical variables	178
10.3	Continuous variables	180
	10.3.1 Fractional polynomials	182
10.4	Interactions	186
10.5	Time-varying variables	189
	10.5.1 Using <code>stcox</code> , <code>tvc()</code> <code>texp()</code>	191

13.6	Generalized gamma regression (AFT metric)	276
13.7	Choosing among parametric models	278
13.7.1	Nested models	278
13.7.2	Nonnested models	281
14	Postestimation commands for parametric models	283
14.1	Use of predict after streg	283
14.1.1	Predicting the time of failure	285
14.1.2	Predicting the hazard and related functions	291
14.1.3	Calculating residuals	294
14.2	Using stcurve	295
15	Generalizing the parametric regression model	301
15.1	Using the ancillary() option	301
15.2	Stratified models	307
15.3	Frailty models	310
15.3.1	Unshared frailty models	311
15.3.2	Example: Kidney data	312
15.3.3	Testing for heterogeneity	317
15.3.4	Shared frailty models	324
16	Power and sample-size determination for survival analysis	333
16.1	Estimating sample size	335
16.1.1	Multiple-myeloma data	336
16.1.2	Comparing two survivor functions nonparametrically	337
16.1.3	Comparing two exponential survivor functions	341
16.1.4	Cox regression models	345
16.2	Accounting for withdrawal and accrual of subjects	348
16.2.1	The effect of withdrawal or loss to follow-up	348
16.2.2	The effect of accrual	349
16.2.3	Examples	351
16.3	Estimating power and effect size	359
16.4	Tabulating or graphing results	360

17	Competing risks	365
17.1	Cause-specific hazards	366
17.2	Cumulative incidence functions	367
17.3	Nonparametric analysis	368
17.3.1	Breast cancer data	369
17.3.2	Cause-specific hazards	369
17.3.3	Cumulative incidence functions	372
17.4	Semiparametric analysis	375
17.4.1	Cause-specific hazards	375
	Simultaneous regressions for cause-specific hazards	378
17.4.2	Cumulative incidence functions	382
	Using <code>stcrreg</code>	382
	Using <code>stcox</code>	389
17.5	Parametric analysis	389
	References	393
	Author index	401
	Subject index	405