

# Contents

List of tables	xiii
List of figures	xv
Preface to the Revised Third Edition	xix
Preface to the Third Edition	xxi
Preface to the Second Edition	xxiii
Preface to the Revised Edition	xxv
Preface to the First Edition	xxvii
Notation and typography	xxix
<b>1 The problem of survival analysis</b>	<b>1</b>
1.1 Parametric modeling . . . . .	2
1.2 Semiparametric modeling . . . . .	2
1.3 Nonparametric analysis . . . . .	5
1.4 Linking the three approaches . . . . .	5
<b>2 Describing the distribution of failure times</b>	<b>7</b>
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
<b>3 Hazard models</b>	<b>19</b>
3.1 Parametric models . . . . .	20
3.2 Semiparametric models . . . . .	21
3.3 Analysis time (time at risk) . . . . .	24

<b>4</b>	<b>Censoring and truncation</b>	<b>29</b>
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	Right-truncation . . . . .	35
4.2.3	Gaps . . . . .	36
<b>5</b>	<b>Recording survival data</b>	<b>37</b>
5.1	The desired format . . . . .	37
5.2	Other formats . . . . .	40
5.3	Example: Wide-form snapshot data . . . . .	44
<b>6</b>	<b>Using stset</b>	<b>47</b>
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
<b>7</b>	<b>After stset</b>	<b>73</b>
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 stfill . . . . .	81
7.6	Example: Hip-fracture data . . . . .	83
<b>8</b>	<b>Nonparametric analysis</b>	<b>91</b>
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	Gaps . . . . .	99
8.2.5	Relationship to the empirical distribution function . . . . .	99
8.2.6	Other uses of sts list . . . . .	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 . . . . .	118
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 . . . . .	126
8.6.4	Stratified tests . . . . .	126
<b>9</b>	<b>The Cox proportional hazards model</b>	<b>131</b>
9.1	Using stcox . . . . .	132
9.1.1	The Cox model has no intercept . . . . .	133
9.1.2	Interpreting coefficients . . . . .	133
9.1.3	The effect of units on coefficients . . . . .	135
9.1.4	Estimating the baseline cumulative hazard and survivor functions . . . . .	137
9.1.5	Estimating the baseline hazard function . . . . .	141
9.1.6	The effect of units on the baseline functions . . . . .	145
9.2	Likelihood calculations . . . . .	147

9.2.1	No tied failures . . . . .	147
9.2.2	Tied failures . . . . .	150
	The marginal calculation . . . . .	150
	The partial calculation . . . . .	151
	The Breslow approximation . . . . .	152
	The Efron approximation . . . . .	153
9.2.3	Summary . . . . .	153
9.3	Stratified analysis . . . . .	155
9.3.1	Obtaining coefficient estimates . . . . .	155
9.3.2	Obtaining estimates of baseline functions . . . . .	157
9.4	Cox models with shared frailty . . . . .	158
9.4.1	Parameter estimation . . . . .	160
9.4.2	Obtaining estimates of baseline functions . . . . .	164
9.5	Cox models with survey data . . . . .	167
9.5.1	Declaring survey characteristics . . . . .	168
9.5.2	Fitting a Cox model with survey data . . . . .	169
9.5.3	Some caveats of analyzing survival data from complex survey designs . . . . .	171
9.6	Cox model with missing data—multiple imputation . . . . .	172
9.6.1	Imputing missing values . . . . .	174
9.6.2	Multiple-imputation inference . . . . .	176
<b>10</b>	<b>Model building using <code>stcox</code></b>	<b>179</b>
10.1	Indicator variables . . . . .	179
10.2	Categorical variables . . . . .	180
10.3	Continuous variables . . . . .	182
	10.3.1 Fractional polynomials . . . . .	184
10.4	Interactions . . . . .	188
10.5	Time-varying variables . . . . .	191
	10.5.1 Using <code>stcox</code> , <code>tvc()</code> <code>texp()</code> . . . . .	193
	10.5.2 Using <code>stsplot</code> . . . . .	195

10.6	Modeling group effects: fixed-effects, random-effects, stratification, and clustering . . . . .	199
<b>11</b>	<b>The Cox model: Diagnostics</b>	<b>205</b>
11.1	Testing the proportional-hazards assumption . . . . .	205
11.1.1	Tests based on reestimation . . . . .	205
11.1.2	Test based on Schoenfeld residuals . . . . .	209
11.1.3	Graphical methods . . . . .	211
11.2	Residuals and diagnostic measures . . . . .	214
	Reye's syndrome data . . . . .	215
11.2.1	Determining functional form . . . . .	216
11.2.2	Goodness of fit . . . . .	221
11.2.3	Outliers and influential points . . . . .	225
<b>12</b>	<b>Parametric models</b>	<b>231</b>
12.1	Motivation . . . . .	231
12.2	Classes of parametric models . . . . .	234
12.2.1	Parametric proportional hazards models . . . . .	236
12.2.2	Accelerated failure-time models . . . . .	241
12.2.3	Comparing the two parameterizations . . . . .	243
<b>13</b>	<b>A survey of basic parametric regression models in Stata</b>	<b>247</b>
13.1	The exponential model . . . . .	249
13.1.1	Exponential regression in the PH metric . . . . .	249
13.1.2	Exponential regression in the AFT metric . . . . .	256
13.2	Weibull regression . . . . .	259
13.2.1	Weibull regression in the PH metric . . . . .	259
	Fitting null models . . . . .	263
13.2.2	Weibull regression in the AFT metric . . . . .	267
13.3	Gompertz regression (PH metric) . . . . .	268
13.4	Lognormal regression (AFT metric) . . . . .	271
13.5	Loglogistic regression (AFT metric) . . . . .	275
13.6	Generalized gamma regression (AFT metric) . . . . .	278

13.7	Choosing among parametric models . . . . .	280
13.7.1	Nested models . . . . .	280
13.7.2	Nonnested models . . . . .	283
<b>14</b>	<b>Postestimation commands for parametric models</b>	<b>285</b>
14.1	Use of predict after streg . . . . .	285
14.1.1	Predicting the time of failure . . . . .	287
14.1.2	Predicting the hazard and related functions . . . . .	293
14.1.3	Calculating residuals . . . . .	296
14.2	Using stcurve . . . . .	298
14.3	Predictive margins and marginal effects . . . . .	302
14.3.1	Predictive margins . . . . .	302
	Marginal mean survival time . . . . .	302
	Marginal survival probabilities . . . . .	307
	Multiple-record data . . . . .	309
14.3.2	Marginal effects . . . . .	311
<b>15</b>	<b>Generalizing the parametric regression model</b>	<b>317</b>
15.0.3	Using the ancillary() option . . . . .	317
15.0.4	Stratified models . . . . .	323
15.1	Frailty models . . . . .	326
15.1.1	Unshared-frailty models . . . . .	327
15.1.2	Example: Kidney data . . . . .	328
15.1.3	Testing for heterogeneity . . . . .	333
15.1.4	Shared-frailty models . . . . .	341
<b>16</b>	<b>Power and sample-size determination for survival analysis</b>	<b>349</b>
16.1	Estimating sample size . . . . .	352
16.1.1	Multiple-myeloma data . . . . .	352
16.1.2	Comparing two survivor functions nonparametrically . . . . .	354
16.1.3	Comparing two exponential survivor functions . . . . .	358
16.1.4	Cox regression models . . . . .	362
16.2	Accounting for withdrawal and accrual of subjects . . . . .	365

16.2.1	The effect of withdrawal or loss to follow-up . . . . .	365
16.2.2	The effect of accrual . . . . .	366
16.2.3	Examples . . . . .	369
16.3	Estimating power and effect size . . . . .	376
16.4	Tabulating or graphing results . . . . .	377
<b>17</b>	<b>Competing risks</b>	<b>381</b>
17.1	Cause-specific hazards . . . . .	382
17.2	Cumulative incidence functions . . . . .	383
17.3	Nonparametric analysis . . . . .	384
17.3.1	Breast cancer data . . . . .	385
17.3.2	Cause-specific hazards . . . . .	386
17.3.3	Cumulative incidence functions . . . . .	389
17.4	Semiparametric analysis . . . . .	392
17.4.1	Cause-specific hazards . . . . .	392
	Simultaneous regressions for cause-specific hazards . . . . .	395
17.4.2	Cumulative incidence functions . . . . .	399
	Using <code>stcrreg</code> . . . . .	399
	Using <code>stcox</code> . . . . .	406
17.5	Parametric analysis . . . . .	407
	<b>References</b>	<b>409</b>
	<b>Author index</b>	<b>417</b>
	<b>Subject index</b>	<b>421</b>