

Contents

Preface to the revised edition	xv
Preface	xvii
Notation and Typography	xxi
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	11
2.3 Interpreting the hazard and cumulative hazard	13
2.3.1 Interpreting the cumulative hazard	14
2.3.2 Interpreting the hazard rate	16
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) 35
 - 4.2.2 Interval truncation (gaps) 36
 - 4.2.3 Right truncation 36
- 5 Recording survival data 37**
 - 5.1 The desired format 37
 - 5.2 Other formats 41
 - 5.3 Example 45
- 6 Using stset 47**
 - 6.1 A short lesson on dates 48
 - 6.2 The purpose of the stset command 50
 - 6.3 The 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 66
 - 6.3.8 Convenience options 69
- 7 After stset 73**
 - 7.1 Look at stset’s output 73
 - 7.2 List some of your data 76
 - 7.3 Use stdes 77
 - 7.4 Use stvary 78

<i>Contents</i>	vii
7.5 Perhaps use stfill	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 sts list	101
8.2.7 Graphing the Kaplan–Meier estimate	102
8.3 The Nelson–Aalen estimator	105
8.4 Estimating the hazard function	111
8.5 Tests of hypothesis	113
8.5.1 The log-rank test	113
8.5.2 The Wilcoxon test	116
8.5.3 Other tests	116
8.5.4 Stratified tests	117
9 The Cox proportional hazards model	121
9.1 Using stcox	122
9.1.1 The Cox model has no intercept	123
9.1.2 Interpreting coefficients	123
9.1.3 The effect of units on coefficients	125
9.1.4 Estimating the baseline cumulative hazard and survivor functions	127
9.1.5 Estimating the baseline hazard function	131
9.1.6 The effect of units on the baseline functions	134
9.2 Likelihood calculations	136
9.2.1 No tied failures	137

- 9.2.2 Tied failures 140
 - The marginal calculation 140
 - The partial calculation 141
 - The Breslow approximation 142
 - The Efron approximation 143
 - 9.2.3 Summary 143
- 9.3 Stratified analysis 144
 - 9.3.1 Obtaining coefficient estimates 144
 - 9.3.2 Obtaining estimates of baseline functions 146
- 9.4 Cox models with shared frailty 147
 - 9.4.1 Parameter estimation 148
 - 9.4.2 Obtaining estimates of baseline functions 152
- 10 Model building using stcox 157**
 - 10.1 Indicator variables 157
 - 10.2 Categorical variables 158
 - 10.3 Continuous variables 160
 - 10.4 Interactions 162
 - 10.5 Time-varying variables 166
 - 10.5.1 Using stcox, tvc() texp() 167
 - 10.5.2 Using stsplint 169
- 11 The Cox model: Diagnostics 175**
 - 11.1 Testing the proportional hazards assumption 175
 - 11.1.1 Tests based on re-estimation 175
 - 11.1.2 Test based on Schoenfeld residuals 178
 - 11.1.3 Graphical methods 181
 - 11.2 Residuals 184
 - Reye’s syndrome data 185
 - 11.2.1 Determining functional form 186
 - 11.2.2 Goodness of fit 190
 - 11.2.3 Outliers and influential points 193

12 Parametric models	197
12.1 Motivation	197
12.2 Classes of parametric models	200
12.2.1 Parametric proportional hazards models	201
12.2.2 Accelerated failure-time models	207
12.2.3 Comparing the two parameterizations	209
13 A survey of parametric regression models in Stata	213
13.1 The exponential model	215
13.1.1 Exponential regression in the PH metric	215
13.1.2 Exponential regression in the AFT metric	222
13.2 Weibull regression	224
13.2.1 Weibull regression in the PH metric	224
Fitting null models	229
13.2.2 Weibull regression in the AFT metric	233
13.3 Gompertz regression (PH metric)	234
13.4 Log-normal regression (AFT metric)	237
13.5 Log-logistic regression (AFT metric)	240
13.6 Generalized gamma regression (AFT metric)	244
13.7 Choosing among parametric models	246
13.7.1 Nested models	246
13.7.2 Non-nested models	249
14 Post-estimation commands for parametric models	251
14.1 Use of predict after streg	251
14.1.1 Predicting the time of failure	253
14.1.2 Predicting the hazard and related functions	259
14.1.3 Calculating residuals	262
14.2 Using stcurve	264

15 Generalizing the parametric regression model	269
15.1 Using the ancillary() option	269
15.2 Stratified models	275
15.3 Frailty models	278
15.3.1 Unshared frailty models	279
15.3.2 Kidney data	280
15.3.3 Testing for heterogeneity	285
15.3.4 Shared frailty models	292
References	299
Author index	303
Subject index	305