

David J. Olive

Survival Analysis

December 11, 2020

Springer

Preface

Many statistics departments offer a one semester undergraduate–graduate course in Reliability and Survival Analysis using texts such as Allison (2010), Collett (2003), and Hosmer et al. (2008). More advanced texts include Harrell (2015), Kalbfleisch and Prentice (2002), Klein and Moeschberger (2003), Lawless (2002), Miller (1981), and Smith (2002). Also see Kleinbaum and Klein (2012), Lee and Wang (2003), Leemis (1995), Meeker and Escobar (1998), and Tableman and Kim (2003).

The prerequisite for this text is a calculus based course in statistics at the level of Chihara and Hesterberg (2011), Hogg and Tanis (2005), Larsen and Marx (2011), Wackerly, Mendenhall and Scheaffer (2008) or Walpole, Myers, Myers and Ye (2006). Linear algebra is essential, and knowledge of regression would be useful. See Olive (2017) and Cook and Weisberg (1999).

Some highlights of this text follow.

- The response plot is useful for checking the model.

Downloading the book’s R functions *survpack.txt* and data files *survdata.txt* into *R*: The commands

```
source("http://parker.ad.siu.edu/Olive/survpack.txt")
source("http://parker.ad.siu.edu/Olive/survdata.txt")
```

The *R* software is used in this text. See R Core Team (2016). Some packages used in the text include

Acknowledgements

Teaching Survival Analysis in Math 473 at Southern Illinois University was very useful.

Contents

1	Univariate Survival Analysis	1
1.1	Functions Related to the Survival Function	1
1.2	Estimating the Survival Function	5
1.3	Estimating the (Cumulative) Hazard Function	15
1.4	Maximum Likelihood Estimation	16
1.5	Simulations for KM Confidence Intervals	20
1.6	Summary	22
1.7	Complements	30
1.8	Problems	30
2	Cox Proportional Hazards Regression	43
2.1	Proportional Hazards Regression	44
2.2	Visualizing the Cox PH Regression Model	46
2.3	Testing	51
2.4	Variable Selection	60
2.5	Stratified Proportional Hazards Regression	65
2.6	Generalized Cox Regression	66
2.7	Summary	67
2.8	Complements	75
2.9	Problems	76
3	Parametric Survival Regression	89
3.1	Weibull and Exponential Regression	89
3.2	Accelerated Failure Time Models	98
3.3	Variable Selection	102
3.4	Summary	102
3.5	Complements	106
3.6	Problems	107

4	Inference After Variable Selection	115
4.1	Variable Selection	115
4.2	Some Tools for Large Sample Theory	116
4.2.1	The Multivariate Normal Distribution	116
4.2.2	The CLT and the Delta Method	120
4.2.3	Modes of Convergence and Consistency	123
4.2.4	Slutsky's Theorem and Related Results	130
4.2.5	Multivariate Limit Theorems	133
4.3	Mixture Distributions	137
4.4	Large Sample Theory for Some Variable Selection	
	Estimators	139
4.5	Prediction Intervals	142
4.6	Prediction Regions	146
4.7	Bootstrapping Hypothesis Tests and Confidence	
	Regions	153
4.7.1	The Bootstrap	155
4.7.2	Bootstrap Confidence Regions for Hypothesis	
	Testing	157
4.7.3	Theory for Bootstrap Confidence Regions	161
4.8	Bootstrapping Variable Selection	166
4.8.1	The Parametric Bootstrap	167
4.8.2	The Nonparametric Bootstrap	168
4.8.3	Bootstrapping Variable Selection	168
4.8.4	Simulations	171
4.9	Data Splitting	174
4.10	Summary	174
4.11	Complements	177
4.12	Problems	178
5	Stuff for Students	181
5.1	R	181
5.2	SAS	185
5.3	Hints for Selected Problems	185
5.4	Tables	186
	Index	197