Biostat/Stat 576

Statistical Methods for Survival Analysis

Lecture 1

March 31, 2009

Chapter 1. Introduction

1. A brief history of time-to-event analysis

2. Time-to-event and censoring

3. Life tables

4. Counting processes and martingales

1. A Brief History

- Prior to 1958, Life-table analysis
- Methods development
  - 1958, Kaplan-Meier estimator, *JASA*
  - 1966, (Mantel) Log-rank, *Cancer Chemo. Rep.*
  - 1972, Cox proportional hazards model, *JRSS-B*
- Theory development
  - 1974, Breslow & Crowley, *Ann. Stat.*
  - 1975, Partial likelihood, *Bmka*
  - 1978, Aalen, *Ann. Stat.*
  - 1981, Tsiatis, *Ann. Stat.*
  - 1982, Andersen & Gill, *Ann. Stat.*
  - Semiparametric regression methods

2. Time-to-event and censoring

- Time-to-event
  - time origin/time zero/starting time
  - staggered entry
  - event/endpoint
  - time-scale
  - causal implication
- Censoring
  - right/left/interval/doubly censored
  - Type-I/Type-II/random censoring
  - current status
  - some examples
    * administrative censoring
    * competing risk
    * lost-to-followup

3. Life tables

- Data example

Years	Alive at beginning	Deaths	Censored
0-1	146	27	3
1-2	116	18	10
2-3	88	21	10
3-4	57	9	3
4-5	45	1	3
		76	29
5-6	41	2	11
6-7	28	3	5
7-8	20	1	8
8-9	11	2	1
9-10	8	2	6

- Question: estimate the 5-year mortality rate? $S(5) = \Pr\{T \geq 5\}$

- Naive estimates

  1. 76 deaths/146 individuals=52.1%, $\widehat{S}(5) = 47.9\%$,

  2. 76 deaths/(146-29)=65%, $\widehat{S}(5) = 35\%$

- Problems

- Life-table estimates:

1. Assume censoring occurred at the right end of interval:

t	n	d	w	$q^r = d/n$	$p^r = 1 - q^r$	$\widehat{S}^r = \prod p^r$
0-1	146	27	3	0.185	0.815	0.815
1-2	116	18	10	0.155	0.845	0.689
2-3	88	21	10	0.239	0.761	0.524
3-4	57	9	3	0.158	0.842	0.441
4-5	45	1	3	0.022	0.972	0.432

5-year survival rate estimate=0.432

2. Assume censoring occurred immediately prior to the left end of interval:

t	n	d	w	$q^l = d/(n-w)$	$p^l = 1 - q^l$	$\widehat{S}^l = \prod p^l$
0-1	146	27	3	0.189	0.811	0.811
1-2	116	18	10	0.170	0.830	0.673
2-3	88	21	10	0.269	0.731	0.492
3-4	57	9	3	0.167	0.833	0.410
4-5	45	1	3	0.024	0.977	0.400

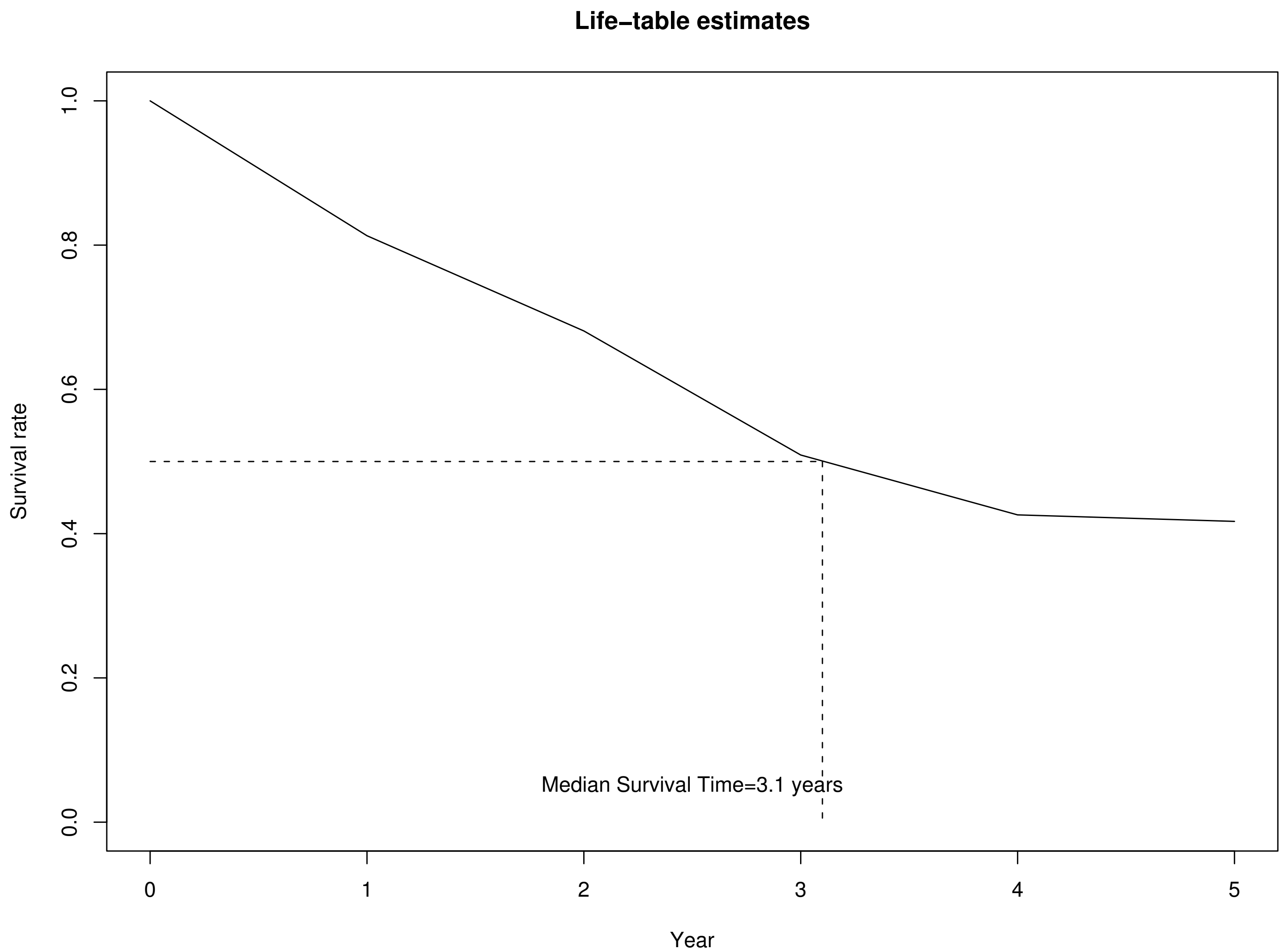
5-year survival rate estimate=0.400

3. Assume half of censoring occurred immediately prior to the left end of interval and half of censoring occurred at the right end of interval:

t	n	d	w	$q = d/(n - w/2)$	$p = 1 - q$	$\hat{S} = \prod p$
0-1	146	27	3	0.187	0.813	0.813
1-2	116	18	10	0.162	0.838	0.681
2-3	88	21	10	0.253	0.747	0.509
3-4	57	9	3	0.162	0.838	0.426
4-5	45	1	3	0.023	0.977	0.417

5-year survival rate estimate=0.417

4. Plot of survival function by life-table method



5. Quick review on Delta-method approximation for variance calculation

(a) $\widehat{\theta} \rightarrow_p \theta$ with known $\mathsf{var}(\widehat{\theta})$

(b) $f(\cdot)$ has sufficient smoothness

(c) First-order Taylor expansion

$$
\begin{aligned}
& f(\widehat{\theta}) - f(\theta) \simeq f'(\theta)(\widehat{\theta} - \theta) \\
\Longrightarrow & E[f(\widehat{\theta}) - f(\theta)]^2 \simeq [f'(\theta)]^2 E(\widehat{\theta} - \theta)^2 \\
\Longrightarrow & \mathsf{var}[f(\widehat{\theta})] + [Ef(\widehat{\theta}) - f(\theta)]^2 \approx [f'(\widehat{\theta})]^2 \mathsf{var}(\widehat{\theta})
\end{aligned}
$$

6. Heuristic justification of variance calculation

(a) $\widehat{S}(t) = \prod p \Longrightarrow \log \widehat{S}(t) = \sum \log p$

(b) $\widehat{\text{var}}(p) = pq/(n - w/2)$

(c) $\widehat{\text{var}}(\log p) \approx \{(\log p)'_p\}^2 \text{var}(p) = q/\{p(n - w/2)\}$

(d) $\text{var}\{\log \widehat{S}(t)\} \approx \sum_{j \le t} d_j/\{(n_j - w_j/2)(n_j - d_j - w_j/2)\}$

(e) $S(t) = \exp\{\log S(t)\} = \exp\{-\Lambda(t)\}$, where $\Lambda(t) = -\log S(t)$

(f) $\text{var}\{\widehat{S}(t)\} = ([\exp\{-\Lambda(t)\}]'_\Lambda)^2 \text{var}\{\widehat{\Lambda}(t)\}$

7. Standard errors for life-table estimates: Greenwood's formula

$$se\left\{\widehat{S}(t)\right\} = \widehat{S}(t)\left\{\sum_{j=1}^{t}\frac{d_j}{(n_j - w_j/2)(n_j - d_j - w_j/2)}\right\}^{1/2}$$

8. 95% confidence intervals for $S(t)$ is $\widehat{S}(t) \pm 1.96 \times se[\widehat{S}(t)]$

9. Life-table example

t	n	d	w	$\widehat{S}$	var(log p)	se
0-1	146	27	3	0.813	0.00159	0.0324
1-2	116	18	10	0.681	0.00174	0.0392
2-3	88	21	10	0.509	0.00408	0.0438
3-4	57	9	3	0.426	0.00349	0.0444
4-5	45	1	3	0.417	0.00054	0.0445

95% confidence interval for $S(5)$: $0.417 \pm 1.96 \times 0.0445 = (0.331, 0.503)$

10. Kaplan-Meier (Product-limit) estimator

(a) subject-level data available

(b) smallest intervals that only one observation occurs in a unit of time

(c) limit of life-tables

(d) nonparametric maximum likelihood estimator