# BIST P8110: Applied Regression II

## 3. Calculating Confidence Intervals for $S(t)$

Qixuan Chen

Department of Biostatistics
Columbia University

# This lecture's big ideas

- Greenwood's variance estimator
- Constructing 95% CI for $S(t)$

# Greenwood's variance estimator

- The variance for $\hat{S}(t)$ was derived by Greenwood:

$$\widehat{Var}\{\hat{S}(t)\} = \hat{S}(t)^2 \left\{ \sum_{j:t_j \leq t} \frac{d_j}{n_j(n_j - d_j)} \right\}$$

# 95% CI for $S(t)$

- What is the potential problem with the following CI estimation?

$$\hat{S}(t) \pm 1.96\sqrt{\widehat{Var}\{\hat{S}(t)\}}$$

# 95% CI for $S(t)$

- What is the potential problem with the following CI estimation?

$$\hat{S}(t) \pm 1.96\sqrt{\widehat{Var}\{\hat{S}(t)\}}$$

- The lower bound can be negative or the upper bound can be greater than 1.

# CI for OR

- What can be an alternative approach to construct 95% CI?
- For odds ratio (OR), a log-transformation is used
  - estimate CI for log-OR
  - back-transform it to obtain CI for OR

# Log-log transformation of $S(t)$

- log-log survival function
  - transform $S(t)$ to $\log[-\log\{S(t)\}]$
  - range changes from $(0, 1)$ to $(-\infty, \infty)$

# Log-log survival function

- Use log-log transformation to construct 95% confidence interval (CI) for $S(t)$
  - compute 95% CI for $\log[-\log\{S(t)\}]$
  - transform back to find 95% CI for $S(t)$
  - the resulted 95% CI for $S(t)$ is guaranteed to lie in $(0, 1)$

# 95% CI for $\log(-\log\{S\})$

- The variance of $\log[-\log\{\hat{S}(t)\}]$ is estimated as

$$\hat{\sigma}^2 = \frac{\widehat{Var}\{\hat{S}(t)\}}{\hat{S}(t)^2[\log \hat{S}(t)]^2}$$
$$= \frac{1}{[\log \hat{S}(t)]^2}\left\{\sum_{j:t_j \le t} \frac{d_j}{n_j(n_j - d_j)}\right\}$$

- A 95% CI for $\log[-\log\{S(t)\}]$ is

$$\log[-\log\{\hat{S}(t)\}] \pm 1.96\sqrt{\hat{\sigma}^2}$$

# 95% CI for $S(t)$ using log-log transformation

- Let $\hat{C}_u = \log[-\log\{\hat{S}(t)\}] + 1.96\sqrt{\hat{\sigma}^2}$
  - the upper bound of the 95% CI for $\log(-\log\{S(t)\})$
- Let $\hat{C}_l = \log[-\log\{\hat{S}(t)\}] - 1.96\sqrt{\hat{\sigma}^2}$
  - the lower bound of the 95% CI for $\log(-\log\{S(t)\})$
- The 95% CI for $S(t)$ is

$$\Big(\exp(-\exp(\hat{C}_u)), \exp(-\exp(\hat{C}_l))\Big)$$

# Example

We continue the cancer recurrence example, suppose we have 8 subjects, among which 5 have cancer returned at 10, 13, 14, 17, and 23 weeks, respectively after treatment. The rest 3 subjects dropped out at 13, 19, and 25 weeks: $10, 13, 13^+, 14, 17, 19^+, 23, 25^+$

| $t_j$ | Interval | $n_j$ | $d_j$ | $p_j = \frac{n_j - d_j}{n_j}$ | $\hat{S}(t_j)$ | 95% CI |
|---|---|---|---|---|---|---|
| 0 | $[0, 10)$ | 8 | 0 | 1 | 1 | |
| 10 | $[10, 13)$ | 8 | 1 | $7/8$ | 0.875 | (0.387, 0.981) |
| 13 | $[13, 14)$ | 7 | 1 | $6/7$ | 0.75 | (0.315, 0.931) |
| 14 | $[14, 17)$ | 5 | 1 | $4/5$ | 0.6 | (0.196, 0.852) |
| 17 | $[17, 23)$ | 4 | 1 | $3/4$ | 0.45 | (0.108, 0.751) |
| 23 | $[23, \infty)$ | 2 | 1 | $1/2$ | 0.225 | (0.012, 0.602) |

# Example (Cont.)

Calculate 95% CI for $S(t)$ with $13 \leq t < 14$

- Step 1: Calculate $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{1}{[\log \hat{S}(t)]^2} \left\{ \sum_{j:t_j \leq t} \frac{d_j}{n_j(n_j - d_j)} \right\} = \frac{1}{[\log(0.75)]^2} \left\{ \frac{1}{8(8-1)} + \frac{1}{7(7-1)} \right\}$$

$$= 0.5035$$

- Step 2: Calculate 95% CI for $\log[-\log\{S(t)\}]$

$$\begin{aligned}(\hat{C}_l, \hat{C}_u) &= \log(-\log\{\hat{S}(t)\}) \pm 1.96 \times \sqrt{\hat{\sigma}^2} \\ &= \log(-\log\{0.75\}) \pm 1.96 \times \sqrt{0.5035} \\ &= (-2.6367, 0.1448)\end{aligned}$$

- Step 3: Calculate 95% CI for S(t)

$$\begin{aligned}&\exp(-\exp(0.1448)), \exp(-\exp(-2.6367)) \\ &= (0.315, 0.931)\end{aligned}$$

# Interpretation of survival functions

- In the cancer recurrence example, we get $\hat{S}(13) = 0.75$ with 95% CI (0.315, 0.931).
- Interpretation
  - We estimate that 75% (95% CI: 31.5% – 93.1%) of patients will survive for more than 13 weeks without cancer recurrence after treatment.

Suggested Readings:

- Chapter 6.1-6.7 (Dupont)
- Chapter 2.1-2.2 (Hosmer et al.)