Nelson-Aalen Estimator

Dr. Mutua Kilai

Department of Pure and Applied Sciences

Jan-April 2024

Innovative technology for a dynamic world

Kirinyaga University

Definition

- The cumulative hazard function $H(t)$ is

$$H(t) = \int_0^t h(t)dt$$

so that

$$S(t) = \exp[-H(t)]$$

and

$$H(t) = -\log S(t)$$

Cont'd

- If we use $\hat{S}_{KM}(t)$ to estimate $S(t)$ for $y_j \leq y < y_{i+1}$ an estimate of the cumulative hazard function can be computed as

$$\begin{aligned} \hat{H}(t) &= -\log[\hat{S_{KM}}(t)] \\ &= -\log\left[\prod_{i=1}^{j}(1 - \frac{w_i}{r_i})\right] \\ &= -\sum_{i=1}^{j}\log\left(1 - \frac{w_i}{r_i}\right) \end{aligned} \tag{1}$$

- w_i be the number of observed events at time t_i
- r_i be the number of individuals at risk at time t_i

Cont'd

- Using the approximation

$$-\log\left(1-\frac{w_i}{r_i}\right)\approx\frac{w_i}{r_i}$$

- We obtain $H\hat{(}t)$ as

$$\hat{H}(t)=\sum_{i=1}^{j}\frac{w_i}{r_i}$$

which is the Nelson-Aalen estimate of the cumulative hazard function.

Variance of Nelson-Aalen Estimate

- An estimate of the variance of $Var[\hat{H}(t_j)]$ can be computed by

$$\hat{Var}[\hat{H(t_j)}] = \sum_{i=1}^{j} \frac{w_i}{r_i^2}$$

- A $100(1-\alpha)\%$ confidence interval assuming normal approximation is given by

$$\hat{H(t_j)} \pm z_{1-\frac{\alpha}{2}} \sqrt{\hat{Var}[\hat{H}(t_j)]}$$

Example

Use the data to compute the Nelson-Aalen Estimate for the Cumulative hazard function.

$$10, 7, 32+, 23, 22, 6, 16, 34+, 32+, 25+, 11+, 20+, 19+, 6, 17+, 35+, 6,$$

$$13, 9+, 6+, 10+$$

Solution

time	n.risk(r_i)	n.event(w_i)	$\frac{w_i}{r_i}$	$\sum \frac{w_i}{r_i}$
6	21	3	0.142	0.142
7	17	1	0.058	0.200
10	15	1	0.067	0.267
13	12	1	0.083	0.350
16	11	1	0.090	0.440
22	7	1	0.142	0.582
23	6	1	0.167	0.749

In R

```r
# reading in the data
data <- c(10,7,32,23,22,6,16,34,32,
          25,11,20,19,6,17,35,6,13,9,6,10)
# status
status <- c(1,1,0,1,1,1,1,0,0,0,0,0,0,
            1,0,0,1,1,0,0,0)
library(survival)
fit <- coxph(Surv(data,status)~1)
tmp1 <- summary(survfit(fit,type="aalen"))
list(tmp1$time,-log(tmp1$surv))
```

Thank You!