MODERN EPIDEMIOLOGY

ASSIGNMENT 1

Hephaes Chuen Chau (M116 I54105029)

2022-03-16

Contents

Questions are adapted from *Epidemiology, Beyond the Basics*.

Question 1

1(a)

We import the data as follows:

```
follow_up_time <- c(2,4,7,8,12,15,17,19,20,23)
event <- c("death","censored","censored","death","censored","death",
           "death","death","censored","death")
dataset <- data.frame(follow_up_time,event)
dataset
```

```
##    follow_up_time    event
## 1               2    death
## 2               4 censored
## 3               7 censored
## 4               8    death
## 5              12 censored
## 6              15    death
## 7              17    death
## 8              19    death
## 9              20 censored
## 10             23    death
```

Using this dataset, we estimate the probability of death at a time i by the following formula:

$$\mathbb{P}(\text{death at time } i) = \frac{\text{number of deaths at } i}{\text{number of at-risk individuals at time } i}$$

while the probability of survival beyond time i is estimated simply to be

$$\mathbb{P}(\text{survival beyond time } i) = 1 - \mathbb{P}(\text{death at time } i)$$

Cumulative probabilities of survival at time j is thus estimated by:

$$\text{Cumulative probabilities of survival at time } j = \prod_{i \leq j} \mathbb{P}(\text{survival beyond time } i)$$

We apply the following function:

```
# function for calculating survival prob
Calculate_P <- function(x = 0, y = 0){
  # x is the counter of loop
  # y is counting the number of dead individuals
  prob_death <<- NA # prob_death is the probability of death at time i
  function(event) {
    if (event == "death") {
      x <<- x + 1
      y <<- y + 1
      message("iteration: ",x)
      message("accumulated number of dead: ",y)
      prob_death <<- 1/(10-x+1)
      return(prob_death)
    } else {
      x <<- x + 1
      message("iteration: ",x)
      message("accumulated number of dead: ",y)
      return(prob_death)
    }
  }
}

# function for calculating cumulative survival prob
Calculate_SP <- function() {
  x <<- 1
  function(y) {
    z <<- x
    x <<- z*y
    return(z*y)
  }
}

# higher-order function initialisation
calculate_p <- Calculate_P()
calculate_sp <- Calculate_SP()

# calculations
prob_death <- sapply(dataset$event, calculate_p)
prob_surv <- 1 - prob_death
cum_prob_surv <- sapply(prob_surv, calculate_sp)

dataset$prob_death <- prob_death
dataset$prob_surv <- prob_surv
dataset$cum_prob_surv <- cum_prob_surv
```

The required answer is therefore

```
dataset
```

```
##    follow_up_time    event prob_death prob_surv cum_prob_surv
```

```
## 1               2    death  0.1000000 0.9000000     0.9000000
## 2               4 censored  0.1000000 0.9000000     0.8100000
## 3               7 censored  0.1000000 0.9000000     0.7290000
## 4               8    death  0.1428571 0.8571429     0.6248571
## 5              12 censored  0.1428571 0.8571429     0.5355918
## 6              15    death  0.2000000 0.8000000     0.4284735
## 7              17    death  0.2500000 0.7500000     0.3213551
## 8              19    death  0.3333333 0.6666667     0.2142367
## 9              20 censored  0.3333333 0.6666667     0.1428245
## 10             23    death  1.0000000 0.0000000     0.0000000
```

where `prob_death` is the probability of death, `prob_surv` is the probability of survival, and `cum_prob_surv` is the cumulative probability of survival.

1(b)

The cumulative probability of survival at the end of the follow-up period is 0, as shown by the table in the previous answer.

1(c)

The cumulative survival probabilities is plotted as below:

```
library(ggplot2)
ggplot(dataset, aes(x=follow_up_time, y=cum_prob_surv)) + geom_step() + theme_classic()
```