Surivival Analysis HW Chapter 2

Victor Feagins

9/8/2021

## Exercise 2.1

Consider the survival data given in Exercise Table 2.1 and compute and plot the **estimated survivorship**, the **probability density** and the **hazard** functions.

| Initial\_Time | End\_Time | Number\_Survivors\_Initial | Number\_Dying\_in\_Interval |
| --- | --- | --- | --- |
| 0 | 1 | 1,100 | 240 |
| 1 | 2 | 860 | 180 |
| 2 | 3 | 680 | 184 |
| 3 | 4 | 496 | 138 |
| 4 | 5 | 358 | 118 |
| 5 | 6 | 240 | 60 |
| 6 | 7 | 180 | 52 |
| 7 | 8 | 128 | 44 |
| 8 | 9 | 84 | 32 |
| 9 |  | 52 | 28 |

### Survivorship Function

Survivorship is the probability that an individual survives longer than t. It is estimated with the following equation:

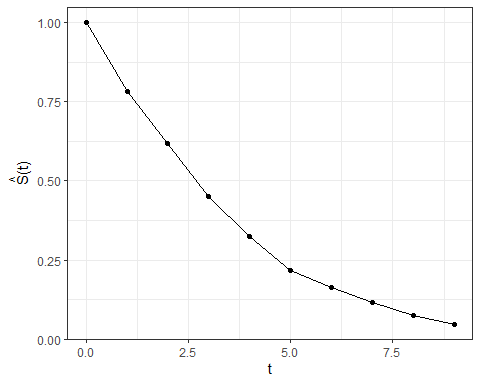
#### Calcuating Survivorship Exercise 2.1

data.2.1 <- data.2.1 %>%   
 mutate('S(t)' = Number\_Survivors\_Initial/max(Number\_Survivors\_Initial))  
data.2.1 %>%   
 select(Initial\_Time, "S(t)") %>%   
 flextable() %>%   
 autofit()

| Initial\_Time | S(t) |
| --- | --- |
| 0 | 1.00000000 |
| 1 | 0.78181818 |
| 2 | 0.61818182 |
| 3 | 0.45090909 |
| 4 | 0.32545455 |
| 5 | 0.21818182 |
| 6 | 0.16363636 |
| 7 | 0.11636364 |
| 8 | 0.07636364 |
| 9 | 0.04727273 |

#### Plotting Survivorship Exercise 2.1

ggplot(data.2.1, aes(`Initial\_Time`, `S(t)`)) +  
 geom\_line() +  
 geom\_point() +  
 ylab(expression(paste(hat("S"), "(t)"))) +  
 xlab("t")



### Probability Density

#### Calcuating probability density Exercise 2.1

To estimate the probability density the following equation is used:

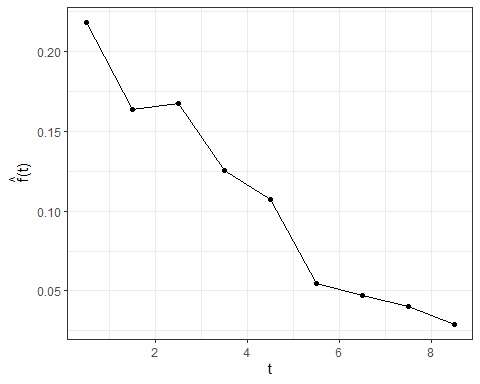
This probability density is at the mid point of the interval

data.2.1 <- data.2.1 %>%   
 mutate(Mid\_Time = (Initial\_Time + End\_Time)/2) %>%   
 mutate("f(t)" = Number\_Dying\_in\_Interval/(max(Number\_Survivors\_Initial) \* (End\_Time - Initial\_Time)))  
data.2.1 %>%   
 select(Mid\_Time, "f(t)") %>%   
 flextable() %>%   
 autofit()

| Mid\_Time | f(t) |
| --- | --- |
| 0.5 | 0.21818182 |
| 1.5 | 0.16363636 |
| 2.5 | 0.16727273 |
| 3.5 | 0.12545455 |
| 4.5 | 0.10727273 |
| 5.5 | 0.05454545 |
| 6.5 | 0.04727273 |
| 7.5 | 0.04000000 |
| 8.5 | 0.02909091 |
|  |  |

#### Plotting probability density Exercise 2.1

ggplot(data.2.1, aes(`Mid\_Time`, `f(t)`)) +  
 geom\_line() +  
 geom\_point() +  
 ylab(expression(paste(hat("f"), "(t)"))) +  
 xlab("t")



### Hazard Function

#### Calcuating Hazard Function Exercise 2.1

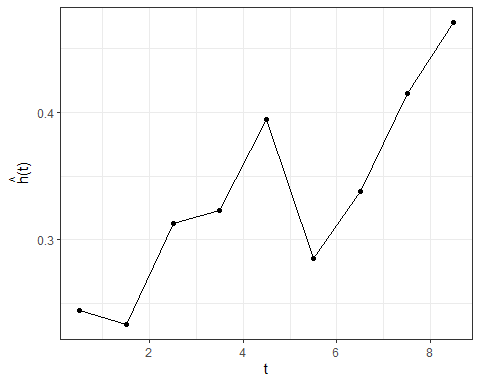
To estimate the hazard function the following equation can be used:

data.2.1 <- data.2.1 %>%   
 mutate(Deaths\_Per\_Time = Number\_Dying\_in\_Interval/(End\_Time - Initial\_Time),  
 "h(t)" = Deaths\_Per\_Time/(Number\_Survivors\_Initial - Number\_Dying\_in\_Interval/2))  
  
data.2.1 %>%   
 select(Mid\_Time, "h(t)") %>%   
 flextable() %>%   
 autofit()

| Mid\_Time | h(t) |
| --- | --- |
| 0.5 | 0.2448980 |
| 1.5 | 0.2337662 |
| 2.5 | 0.3129252 |
| 3.5 | 0.3231850 |
| 4.5 | 0.3946488 |
| 5.5 | 0.2857143 |
| 6.5 | 0.3376623 |
| 7.5 | 0.4150943 |
| 8.5 | 0.4705882 |
|  |  |

#### Plotting Hazard Function Exercise 2.1

ggplot(data.2.1, aes(`Mid\_Time`, `h(t)`)) +  
 geom\_line() +  
 geom\_point() +  
 ylab(expression(paste(hat("h"), "(t)"))) +  
 xlab("t")



## Exercise 2.2

Table 2.2 is a life table for the total population of 100,000 live births in the United States, 1959 – 1961. Compute and plot the estimated **survivorship**, the **probability density**, and the **hazard** function.

| Initial\_Time | End\_Time | Number\_Survivors\_Initial | Number\_Dying\_in\_Interval |
| --- | --- | --- | --- |
| 0 | 1 | 100,000 | 2,593 |
| 1 | 5 | 97,407 | 409 |
| 5 | 10 | 96,998 | 233 |
| 10 | 15 | 96,765 | 214 |
| 15 | 20 | 96,551 | 440 |
| 20 | 25 | 96,111 | 594 |
| 25 | 30 | 95,517 | 612 |
| 30 | 35 | 94,905 | 761 |
| 35 | 40 | 94,144 | 1,080 |
| 40 | 45 | 93,064 | 1,686 |
| 45 | 50 | 91,378 | 2,622 |
| 50 | 55 | 88,756 | 4,045 |
| 55 | 60 | 84,711 | 5,644 |
| 60 | 65 | 79,067 | 7,920 |
| 65 | 70 | 71,147 | 10,290 |
| 70 | 75 | 60,857 | 12,687 |
| 75 | 80 | 48,170 | 14,594 |
| 80 | 85 | 33,576 | 15,034 |
| 85 |  | 18,542 | 18,542 |

### Survivorship Function

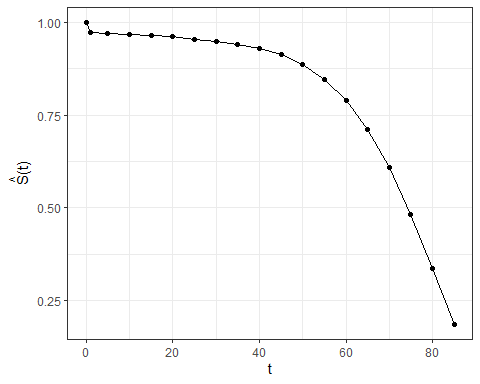
#### Calcuating Survivorship Exercise 2.2

data.2.2 <- data.2.2 %>%   
 mutate('S(t)' = Number\_Survivors\_Initial/max(Number\_Survivors\_Initial))  
data.2.2 %>%   
 select(Initial\_Time, "S(t)") %>%   
 flextable() %>%   
 autofit()

| Initial\_Time | S(t) |
| --- | --- |
| 0 | 1.00000 |
| 1 | 0.97407 |
| 5 | 0.96998 |
| 10 | 0.96765 |
| 15 | 0.96551 |
| 20 | 0.96111 |
| 25 | 0.95517 |
| 30 | 0.94905 |
| 35 | 0.94144 |
| 40 | 0.93064 |
| 45 | 0.91378 |
| 50 | 0.88756 |
| 55 | 0.84711 |
| 60 | 0.79067 |
| 65 | 0.71147 |
| 70 | 0.60857 |
| 75 | 0.48170 |
| 80 | 0.33576 |
| 85 | 0.18542 |

#### Ploting Survivorship Exercise 2.2

ggplot(data.2.2, aes(`Initial\_Time`, `S(t)`)) +  
 geom\_line() +  
 geom\_point() +  
 ylab(expression(paste(hat("S"), "(t)"))) +  
 xlab("t")



### Probability Density

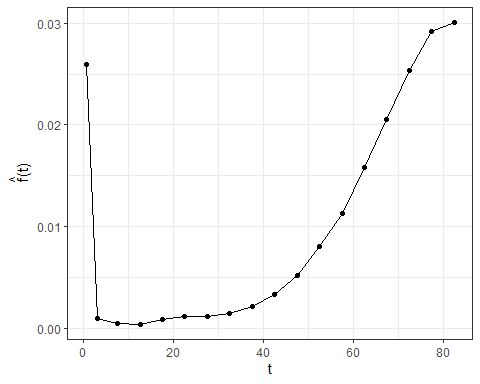
#### Calcuating probability density Exercise 2.2

data.2.2 <- data.2.2 %>%   
 mutate(Mid\_Time = (Initial\_Time + End\_Time)/2) %>%   
 mutate("f(t)" = Number\_Dying\_in\_Interval/(max(Number\_Survivors\_Initial) \* (End\_Time - Initial\_Time)))  
data.2.2 %>%   
 select(Mid\_Time, "f(t)") %>%   
 flextable() %>%   
 autofit()

| Mid\_Time | f(t) |
| --- | --- |
| 0.5 | 0.0259300 |
| 3.0 | 0.0010225 |
| 7.5 | 0.0004660 |
| 12.5 | 0.0004280 |
| 17.5 | 0.0008800 |
| 22.5 | 0.0011880 |
| 27.5 | 0.0012240 |
| 32.5 | 0.0015220 |
| 37.5 | 0.0021600 |
| 42.5 | 0.0033720 |
| 47.5 | 0.0052440 |
| 52.5 | 0.0080900 |
| 57.5 | 0.0112880 |
| 62.5 | 0.0158400 |
| 67.5 | 0.0205800 |
| 72.5 | 0.0253740 |
| 77.5 | 0.0291880 |
| 82.5 | 0.0300680 |
|  |  |

#### Plotting probability density Exercise 2.2

ggplot(data.2.2, aes(`Mid\_Time`, `f(t)`)) +  
 geom\_line() +  
 geom\_point() +  
 ylab(expression(paste(hat("f"), "(t)"))) +  
 xlab("t")



### Hazard Function

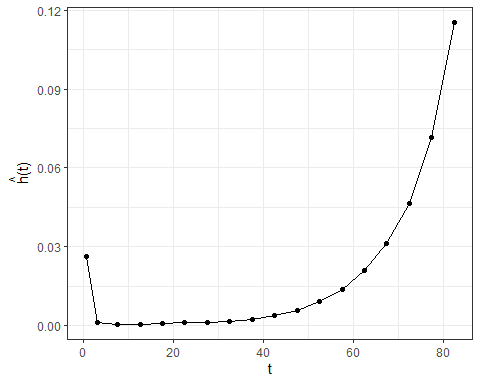
#### Calcuating Hazard Function Exercise 2.2

data.2.2 <- data.2.2 %>%   
 mutate(Deaths\_Per\_Time = Number\_Dying\_in\_Interval/(End\_Time - Initial\_Time),  
 "h(t)" = Deaths\_Per\_Time/(Number\_Survivors\_Initial - Number\_Dying\_in\_Interval/2))  
  
data.2.2 %>%   
 select(Mid\_Time, "h(t)") %>%   
 flextable() %>%   
 autofit()

| Mid\_Time | h(t) |
| --- | --- |
| 0.5 | 0.0262705983 |
| 3.0 | 0.0010519277 |
| 7.5 | 0.0004810000 |
| 12.5 | 0.0004427983 |
| 17.5 | 0.0009135169 |
| 22.5 | 0.0012399023 |
| 27.5 | 0.0012855657 |
| 32.5 | 0.0016101646 |
| 37.5 | 0.0023075937 |
| 42.5 | 0.0036564340 |
| 47.5 | 0.0058223323 |
| 52.5 | 0.0093274225 |
| 57.5 | 0.0137845132 |
| 62.5 | 0.0210899117 |
| 67.5 | 0.0311808733 |
| 72.5 | 0.0465462684 |
| 77.5 | 0.0714114452 |
| 82.5 | 0.1153843202 |
|  |  |

#### Plotting Hazard Function Exercise 2.2

ggplot(data.2.2, aes(`Mid\_Time`, `h(t)`)) +  
 geom\_line() +  
 geom\_point() +  
 ylab(expression(paste(hat("h"), "(t)"))) +  
 xlab("t")



## Exercise 2.5

Given the survivorship function:

derive the probability density function and the hazard function.

### Relationships

We know that the cumulative Hazard function is the negative natural log of the survival function

We know that the cumulative Hazard function is the integral of the hazard function

We also know that the probability density function can be derived as such

### Deriving the cumulative hazard function

### Deriving the hazard function

### Deriving the probability density function