Surivival Analysis HW Chapter 2

Victor Feagins

8/30/2021

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.

Survivorship Function

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

$$\hat{S}(t) = \frac{number\ of\ items\ surviving\ longer\ than\ t}{total\ number\ of\ patients}$$

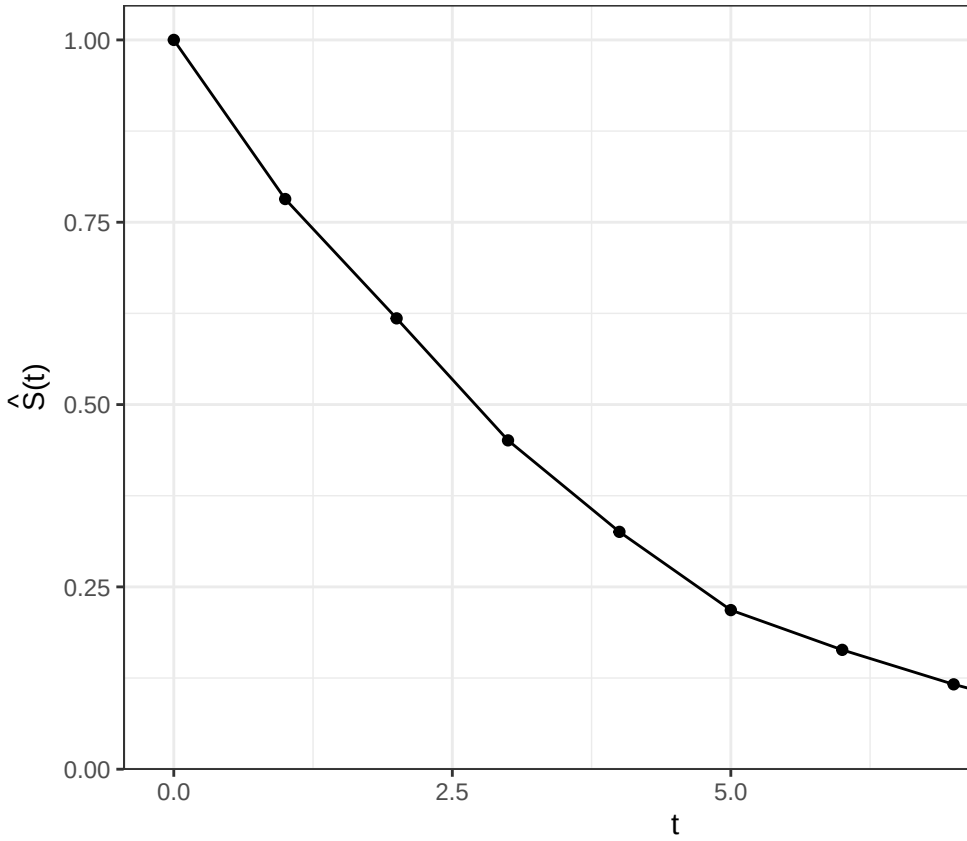
```
data.2.1 <- data.2.1 %>%
  mutate('S(t)' = Number_Survivors_Initial/max(Number_Survivors_Initial))
data.2.1 %>%
  select(Initial_Time, "S(t)") %>%
  kbl() %>%
  kable_classic()
```

Calcuating 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")
```

Initial_Time	End_Time	Number_Survivors_Initial	Number_Dying_in_Interval
0	1	1100	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	NA	52	28

Initial_Time	S(t)
0	1.0000000
1	0.7818182
2	0.6181818
3	0.4509091
4	0.3254545
5	0.2181818
6	0.1636364
7	0.1163636
8	0.0763636
9	0.0472727



Plotting Survivorship Exercise 2.1

Probability Density

Calcuating probability density Exercise 2.1 To estimate the probability density the following equation is used:

$$\hat{f}(t) = \frac{number\ of\ items\ dying\ in\ the\ interval\ beginning\ at\ time\ t}{(total\ number\ of\ items) * (interval\ width)}$$

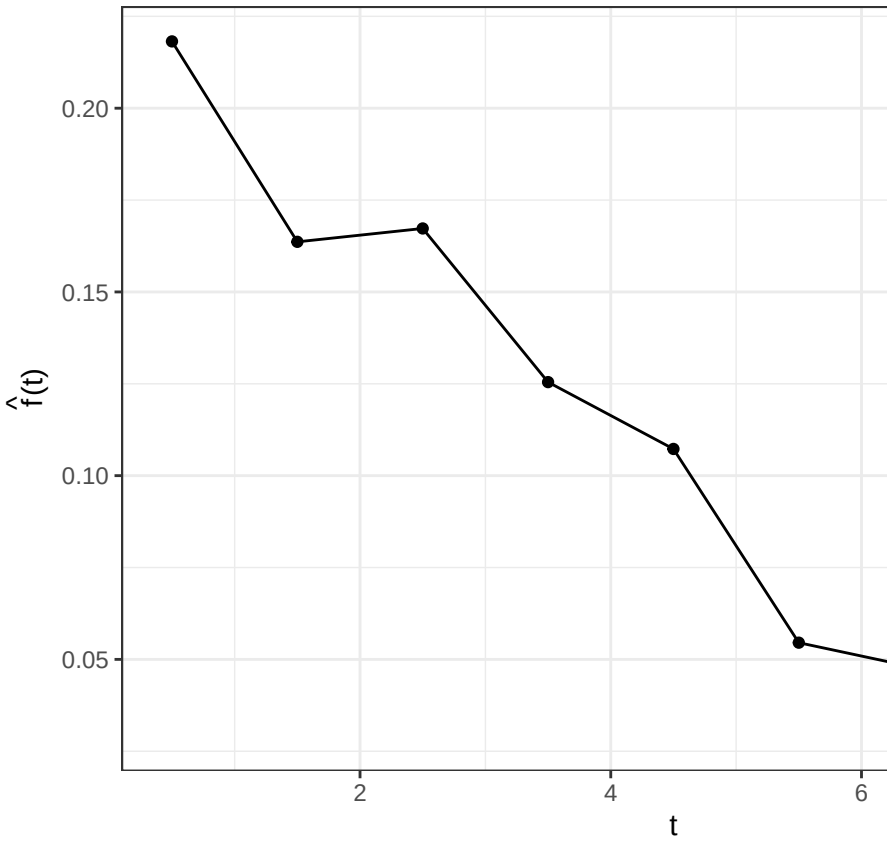
This probability density is at the mid point of the interval

```
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 %>%
```

Mid_Time	f(t)
0.5	0.2181818
1.5	0.1636364
2.5	0.1672727
3.5	0.1254545
4.5	0.1072727
5.5	0.0545455
6.5	0.0472727
7.5	0.0400000
8.5	0.0290909
NA	NA

```
  select(Mid_Time, "f(t)") %>%
  kbl() %>%
  kable_classic()
```

```
ggplot(data.2.1, aes(`Mid_Time`, `f(t)`)) +
  geom_line() +
  geom_point() +
  ylab(expression(paste(hat("f"), "(t)"))) +
  xlab("t")
```



Plotting probability density Exercise 2.1

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
NA	NA

Hazard Function

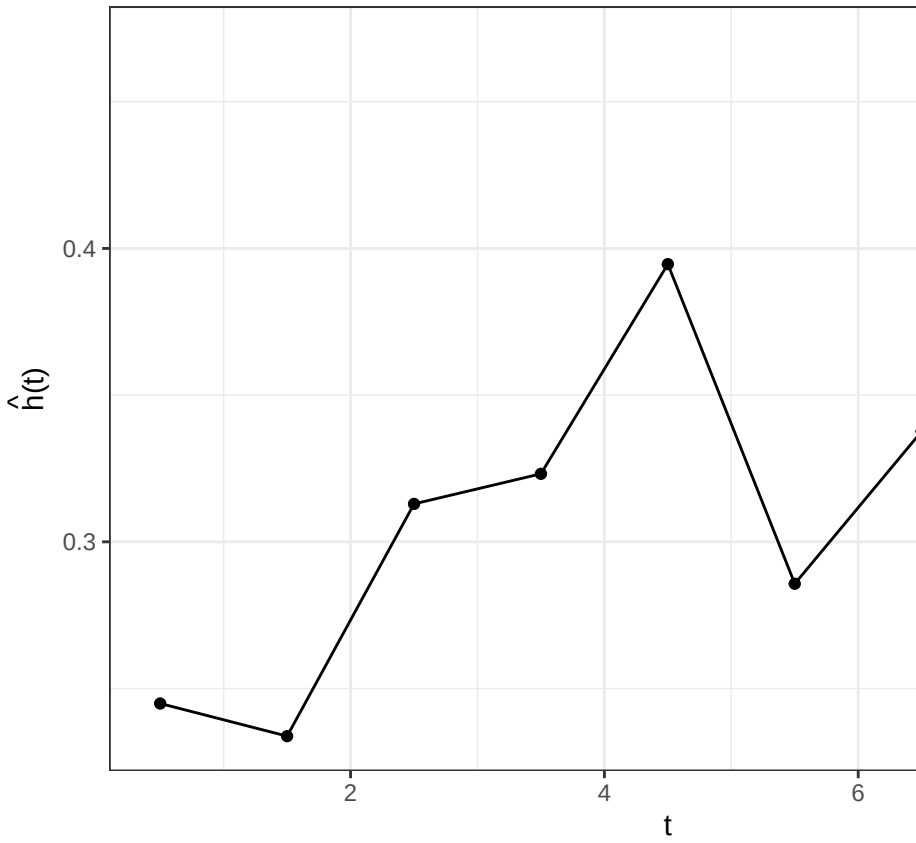
Calcuating Hazard Function Exercise 2.1 To estimate the hazard function the following equation can be used:

$$\hat{h}(t) = \frac{number\ of\ items\ dying\ per\ unit\ time\ in\ the\ interval}{(number\ of\ items\ surviving\ at\ t) - (number\ of\ deaths\ in\ the\ interval)/2}$$

```
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)") %>%
  kbl() %>%
  kable_classic()
```

```
ggplot(data.2.1, aes(`Mid_Time`, `h(t)`)) +
  geom_line() +
  geom_point() +
  ylab(expression(paste(hat("h"), "(t)"))) +
  xlab("t")
```



Plotting Hazard Function Exercise 2.1

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.

Survivorship Function

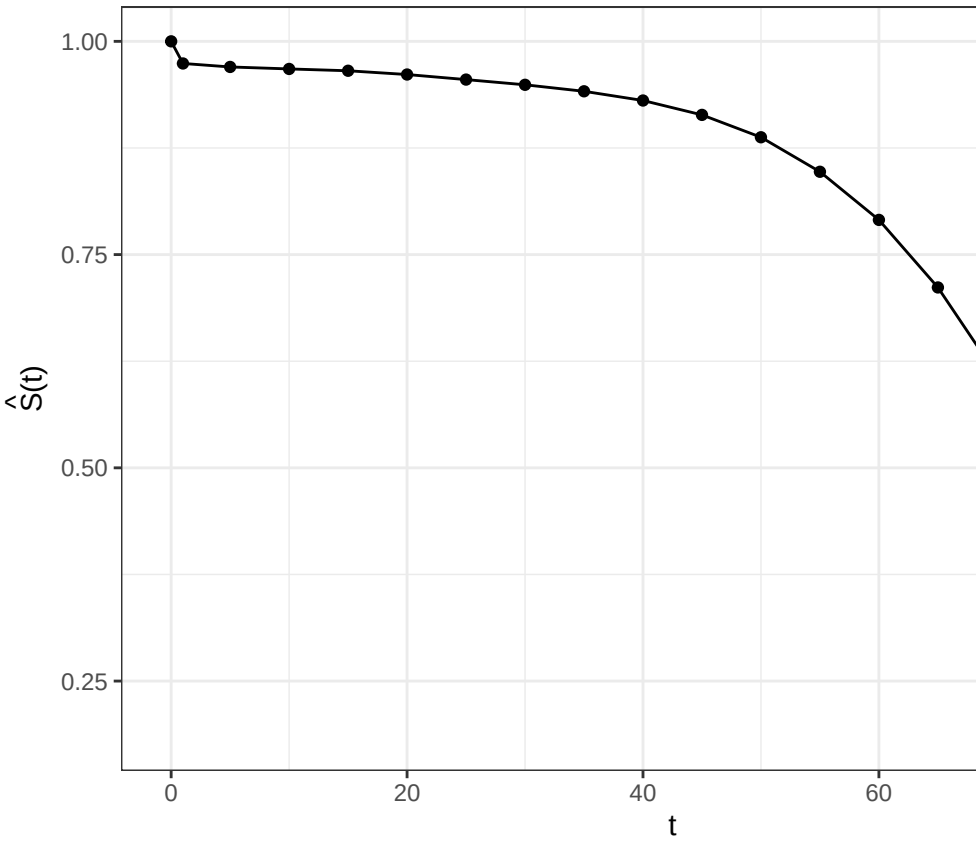
```
data.2.2 <- data.2.2 %>%
  mutate('S(t)' = Number_Survivors_Initial/max(Number_Survivors_Initial))
data.2.2 %>%
  select(Initial_Time, "S(t)") %>%
  kbl() %>%
  kable_classic()
```

Calcuating 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")
```

Initial_Time	End_Time	Number_Survivors_Initial	Number_Dying_in_Interval
0	1	100000	2593
1	5	97407	409
5	10	96998	233
10	15	96765	214
15	20	96551	440
20	25	96111	594
25	30	95517	612
30	35	94905	761
35	40	94144	1080
40	45	93064	1686
45	50	91378	2622
50	55	88756	4045
55	60	84711	5644
60	65	79067	7920
65	70	71147	10290
70	75	60857	12687
75	80	48170	14594
80	85	33576	15034
85	NA	18542	18542

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

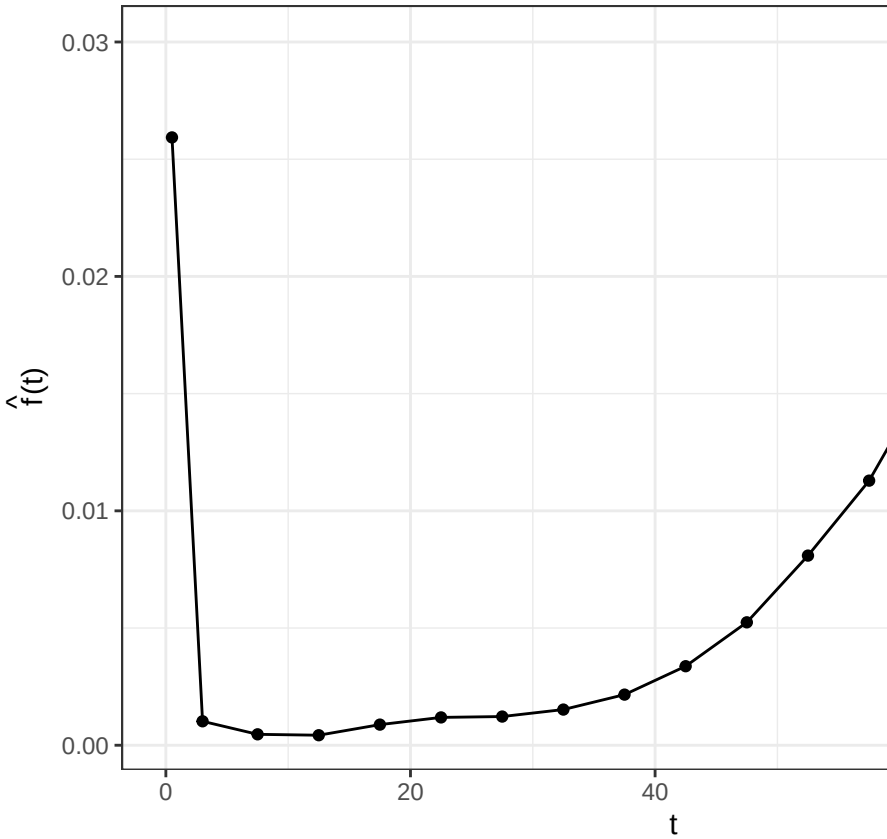
Probability Density

```
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)") %>%
  kbl() %>%
  kable_classic()
```

Calcuating 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")
```

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
NA	NA



Plotting probability density Exercise 2.2

Hazard Function

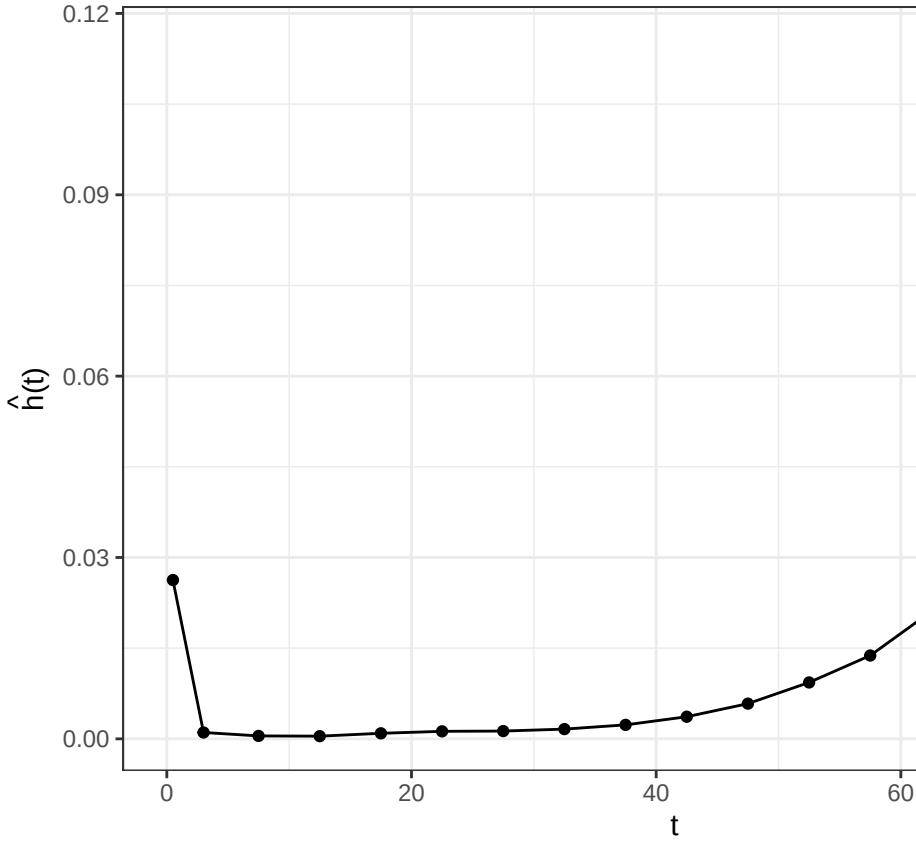
Mid_Time	h(t)
0.5	0.0262706
3.0	0.0010519
7.5	0.0004810
12.5	0.0004428
17.5	0.0009135
22.5	0.0012399
27.5	0.0012856
32.5	0.0016102
37.5	0.0023076
42.5	0.0036564
47.5	0.0058223
52.5	0.0093274
57.5	0.0137845
62.5	0.0210899
67.5	0.0311809
72.5	0.0465463
77.5	0.0714114
82.5	0.1153843
NA	NA

```
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)") %>%
  kbl() %>%
  kable_classic()
```

Calcuating 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")
```



Plotting Hazard Function Exercise 2.2