

MSMS 308 : Practical 01

Ananda Biswas

November 8, 2025

➔ Question

Consider the following survival data of 40 patients with myeloma. Compute and plot the estimated survival function, the probability density function, and the hazard function.

Survival Time t (months)	Number of Patients Surviving at Beginning of the Interval	Number of Patients Dying in the Interval
0 – 5	40	5
5 – 10	35	7
10 – 15	28	6
15 – 20	22	4
20 – 25	18	5
25 – 30	13	4
30 – 35	9	4
35 – 40	5	0
40 – 45	5	2
45 – 50	3	1
≥ 50	2	2

➔ R Program, Plot and Interpretation

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

$$\widehat{f}(t) = \frac{\text{number of patients dying in the interval beginning at time } t}{(\text{total number of patients}) \times (\text{interval width})}$$

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

```
survival_data <- read.csv('https://raw.githubusercontent.com/sakunisgithub/data_sets/refs/heads/master/msc_semester_3/life_time_prac_1.csv',
  stringsAsFactors = TRUE)
```

```
total_number_of_patients <- survival_data$no_at_risk[1]
```

```
S_t_hat <- survival_data$no_at_risk / total_number_of_patients
```

```
interval_width <- 5
```

```
f_t_hat <- survival_data$no_of_death / (total_number_of_patients * interval_width)
```

```
f_t_hat[length(f_t_hat)] = NA
```

```
# a = number_of_patients_dying_per_unit_time_in_the_interval
```

```
a <- survival_data$no_of_death / interval_width
```

```
h_t_hat <- a / (survival_data$no_at_risk - survival_data$no_of_death / 2)
```

```
h_t_hat[length(h_t_hat)] = NA
```

```
analysis_table <- data.frame(survival_data,
  t = (0:10)*5,
  S_t_hat = round(S_t_hat, digits = 3),
  f_t_hat = round(f_t_hat, digits = 3),
  h_t_hat = round(h_t_hat, digits = 3))
```

```
analysis_table
```

##	survival_time	no_at_risk	no_of_death	t	S_t_hat	f_t_hat	h_t_hat
## 1	0--5	40	5	0	1.000	0.025	0.027
## 2	5--10	35	7	5	0.875	0.035	0.044
## 3	10--15	28	6	10	0.700	0.030	0.048
## 4	15--20	22	4	15	0.550	0.020	0.040
## 5	20--25	18	5	20	0.450	0.025	0.065
## 6	25--30	13	4	25	0.325	0.020	0.073
## 7	30--35	9	4	30	0.225	0.020	0.114
## 8	35--40	5	0	35	0.125	0.000	0.000
## 9	40--45	5	2	40	0.125	0.010	0.100
## 10	45--50	3	1	45	0.075	0.005	0.080
## 11	>=50	2	2	50	0.050	NA	NA

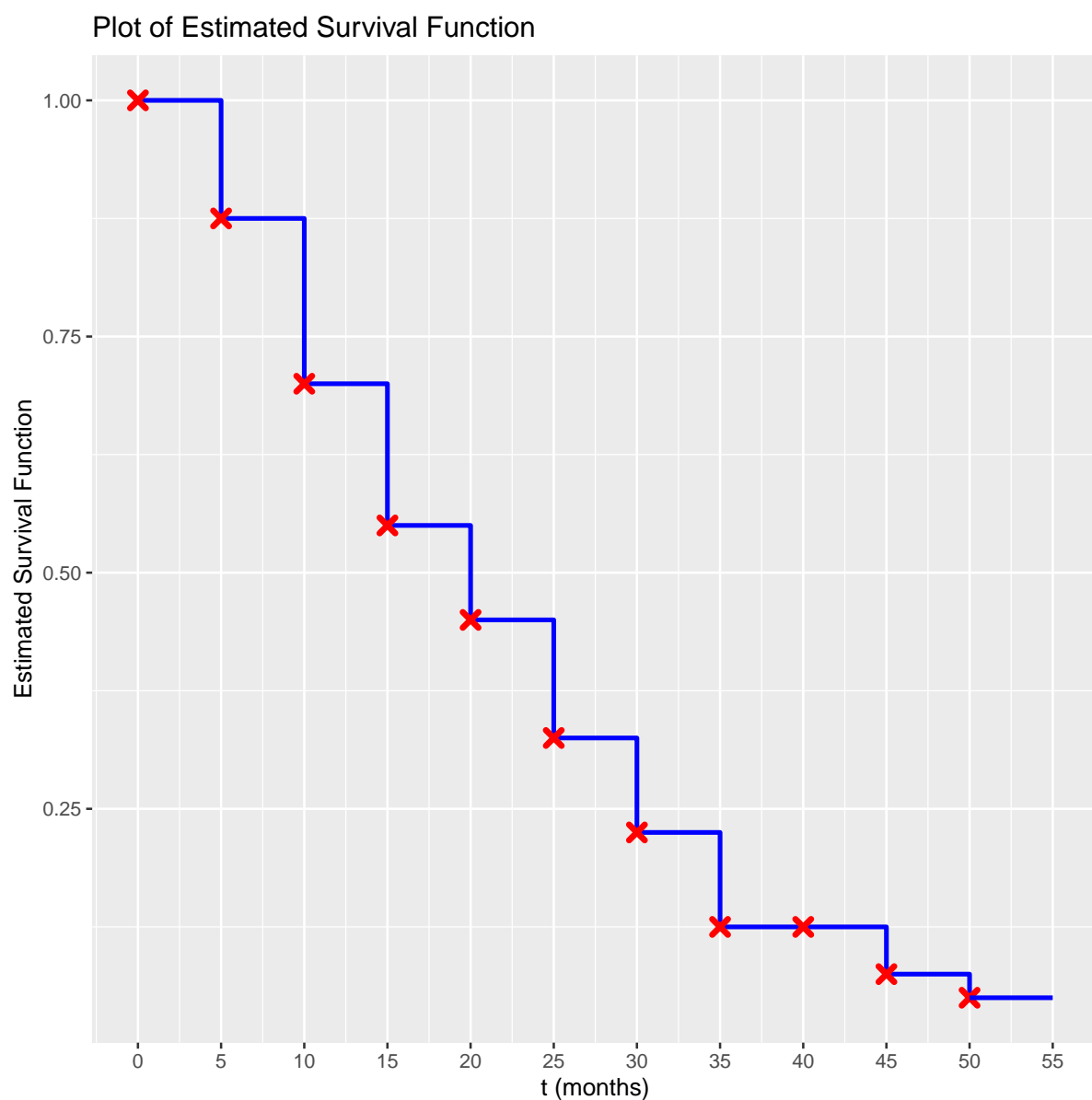
```


t_last <- max(analysis_table$t)

S_t_hat_last <- analysis_table %>% filter(t == t_last) %>% pull(S_t_hat)

analysis_table %>%
  ggplot(aes(x = t, y = S_t_hat)) +
  geom_step(direction = "hv", linewidth = 1, color = "blue") +
  geom_point(size = 2, color = "red", shape = 4, stroke = 2) +
  annotate("segment", x = t_last, xend = t_last + 5,
           y = S_t_hat_last, yend = S_t_hat_last, color = "blue", linewidth = 1) +
  scale_x_continuous(limits = c(0, 55), breaks = seq(0, 55, 5)) +
  labs(x = "t (months)",
       y = "Estimated Survival Function",
       title = "Plot of Estimated Survival Function")

```



 The median survival time for myeloma patients is approximately 17.5 months.

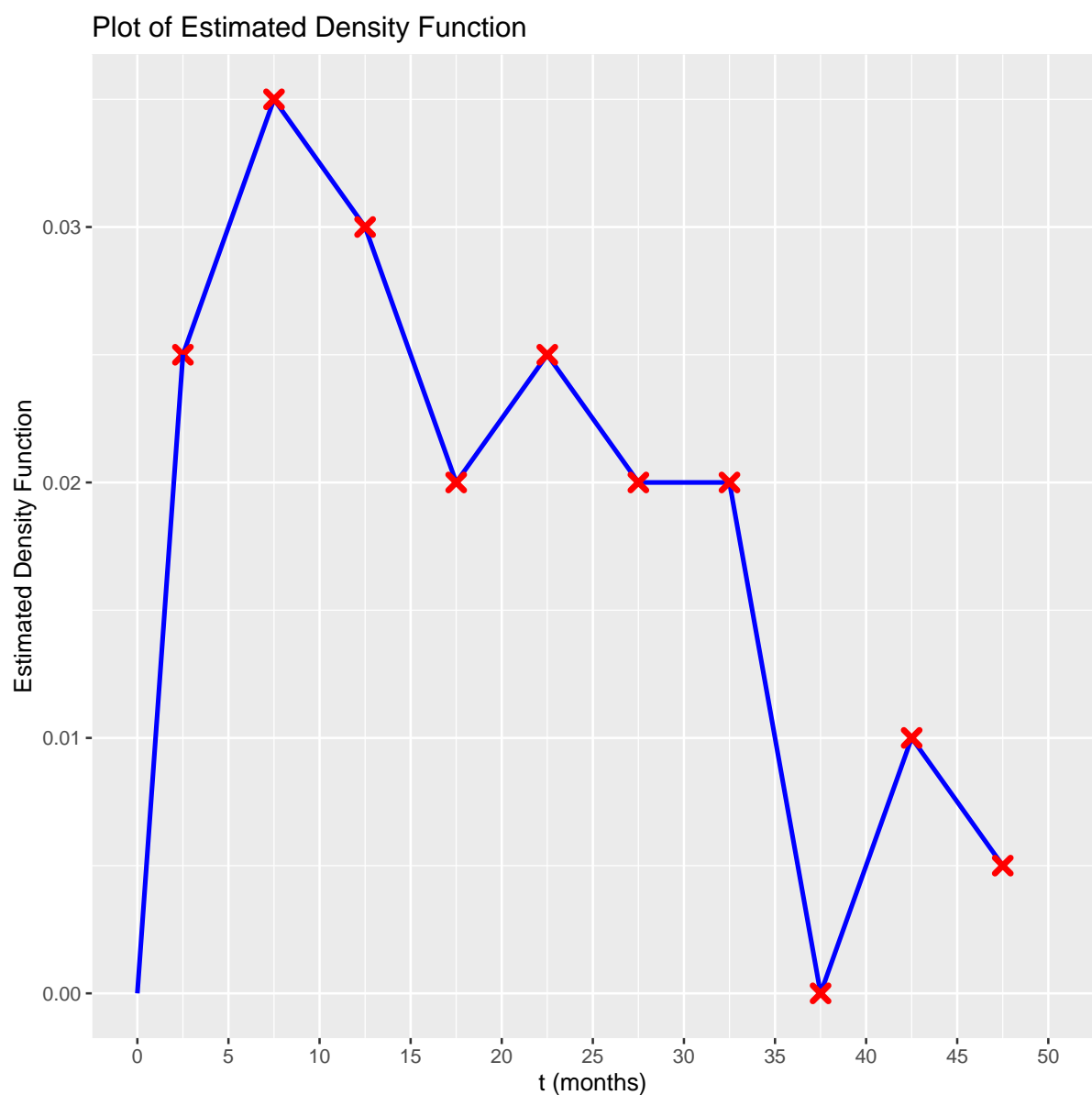
```


t_first <- min(analysis_table$t)

f_t_hat_first <- analysis_table %>% filter(t == t_first) %>% pull(f_t_hat)

analysis_table[-dim(analysis_table)[1],] %>%
  ggplot(aes(x = t+2.5, y = f_t_hat)) +
  geom_line(linewidth = 1, col = "blue") +
  geom_point(size = 2, col = "red", shape = 4, stroke = 2) +
  annotate("segment", x = t_first, xend = t_first + 2.5,
           y = 0, yend = f_t_hat_first, color = "blue", linewidth = 1) +
  scale_x_continuous(limits = c(0, 50), breaks = seq(0, 50, 5)) +
  labs(x = "t (months)",
       y = "Estimated Density Function",
       title = "Plot of Estimated Density Function")

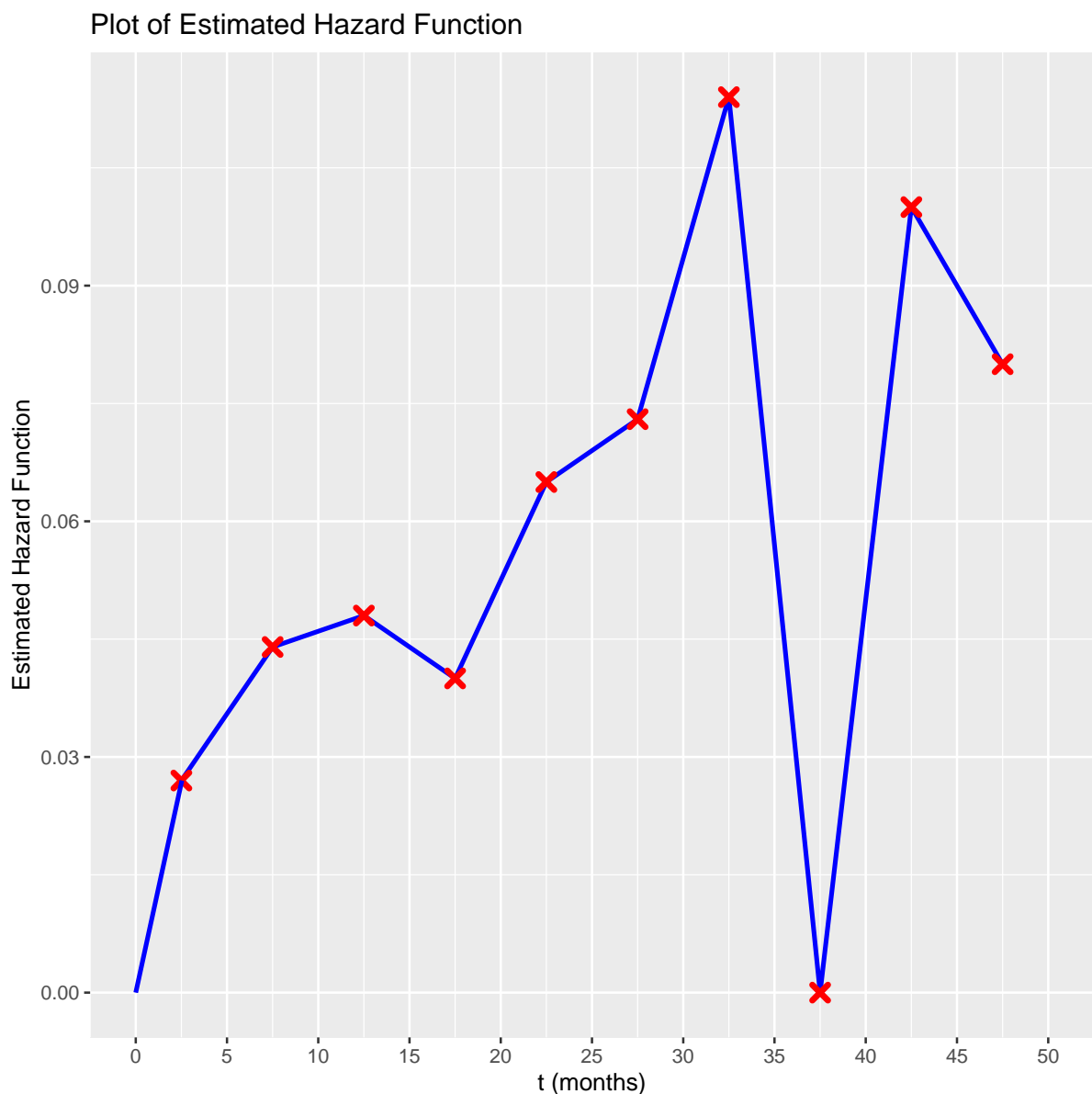
```




 Death due to myeloma is most likely occur in 5 to 10 months.

```
h_t_hat_first = analysis_table %>% filter(t == t_first) %>% pull(h_t_hat)

analysis_table[-dim(analysis_table)[1],] %>%
  ggplot(aes(x = t+2.5, y = h_t_hat)) +
  geom_line(linewidth = 1, col = "blue") +
  geom_point(size = 2, col = "red", shape = 4, stroke = 2) +
  annotate("segment", x = 0, xend = h_t_hat_first + 2.5,
           y = 0, yend = h_t_hat_first, color = "blue", linewidth = 1) +
  scale_x_continuous(limits = c(0, 50), breaks = seq(0, 50, 5)) +
  labs(x = "t (months)",
       y = "Estimated Hazard Function",
       title = "Plot of Estimated Hazard Function")
```



 The hazard function shows an increasing trend and reaches its peak in 30 to 35 months, so risk of death increases over time.