

TMA4275 Lifetime Analysis

Obligatory project 1

Jim Totland

2/4/2022

In this exercise we consider results from an old investigation to evaluate a histochemical marker that discriminates between primary breast cancer that has metastasized and that which has not. The marker under study is denoted HPA. Each tumor was treated with this marker and hence classified as either positively or negatively stained. The data which will be used is given below. The survival times of each woman is given in months and classified according to whether their tumor was negatively or positively stained. Censored survival times are labeled with an asterisk (*).

Negative	Positive
23	5
47	8
69	10
70*	13
71*	18
100*	24
101*	26
148	31
181	35
198*	50
208*	59
212*	61
224*	76*
	109*
	116*
	118
	143
	154*
	162*
	225*

In the following we denote patients with negatively stained tumors as group 1 and patients with positively stained tumors as group 2.

Problem 1

a)

First, a data frame containing the data presented above is constructed.

Next, the risk set for each of the two groups, $Y_1(t)$ and $Y_2(t)$ respectively, are plotted as functions of the study time, t . The result is given in figure 1.

Number of Individuals

t (study time)

Group

- 1
- 2

From figure 1, we observe that group 2 has a much steeper trend than group 1.

b)

Below, a function computing the Nelson-Aalen estimator of the integrated hazard rate, $A(t)$, over the grid t and a corresponding confidence interval with significance level α is given. We refer to the comments in the code for further documentation.

```
N.A.est <- function(df, alpha, conf.int.type, t){

  # df: data frame with survival times
  # alpha: significance level of conf.int
  # conf.int.type: type of confidence interval, (1) regular or (2) log-type.
  # t: time grid

  y <- rep(0, length(t))
  for(i in 1:length(t)){
    y[i] <- sum(df$months > t[i])
  }
  A.hat <- cumsum(1/y)
  # Fix infinite values:
  A.hat[min(which(!is.finite(A.hat))):length(A.hat)] <- A.hat[min(which(!is.finite(A.hat))) - 1]
  sigma.hat <- sqrt(cumsum(1/y^2))
  # Fix infinite values:
  sigma.hat[min(which(!is.finite(sigma.hat))):length(sigma.hat)] <- sigma.hat[min(which(!is.finite(sigma.hat))) - 1]
  z = qnorm(1 - alpha/2)
  upper <- lower <- NA
  if(conf.int.type == 1){
    upper <- A.hat + z*sigma.hat
    lower <- A.hat - z*sigma.hat
  }
  else if(conf.int.type == 2){
    upper <- A.hat * exp(z*sigma.hat/A.hat)
    lower <- A.hat * exp(-z*sigma.hat/A.hat)
  }
  else{
    stop("Invalid conf.int.type")
  }
  return(data.frame(A.hat = A.hat, upper = upper, lower = lower, t = t))
}
```

In the following, the Nelson-Aalen estimator is computed and plotted for the two groups. Figure 2 uses regular confidence intervals, while in figure 3, the confidence interval based on the log-transform is used.

```
df1 <- N.A.est(group1, 0.05, 1, t)
df2 <- N.A.est(group2, 0.05, 1, t)
ggplot(df1) + geom_step(aes(x = t, y = A.hat, color = "1")) +
  geom_stepconfint(aes(x = t, ymin=lower, ymax=upper), fill = "#e0474c", alpha = 0.2) +
  geom_step(data = df2, aes(x = t, y = A.hat, color = "2")) +
  geom_stepconfint(data = df2, aes(x = t, ymin=lower, ymax=upper), fill = "#7ab8d6", alpha = 0.2) +
  scale_color_manual(name = "Group", values = c("1" = "#e0474c", "2" = "#7ab8d6"))
```

```
df1 <- N.A.est(group1, 0.05, 2, t)
df2 <- N.A.est(group2, 0.05, 2, t)
ggplot(df1) + geom_step(aes(x = t, y = A.hat, color = "1")) +
```

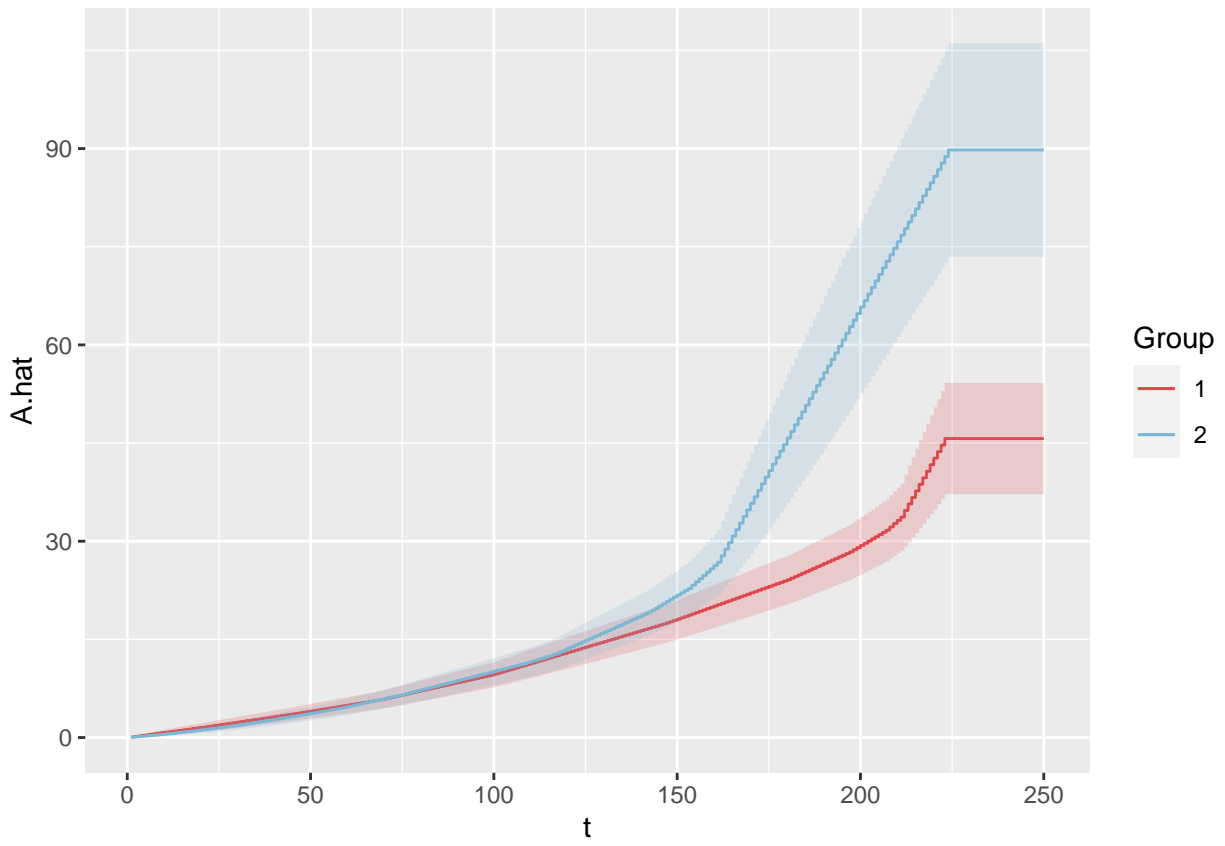


Figure 2: The Nelson-Aalen estimator of the integrated hazard rate with regular confidence intervals.

```
geom_stepconfint(aes(x = t, ymin=lower, ymax=upper), fill = "#e0474c", alpha = 0.2) +
geom_step(data = df2, aes(x = t, y = A.hat, color = "2")) +
geom_stepconfint(data = df2, aes(x = t, ymin=lower, ymax=upper), fill = "#7ab8d6", alpha = 0.2) +
scale_color_manual(name = "Group", values = c("1" = "#e0474c", "2" = "#7ab8d6"))
```

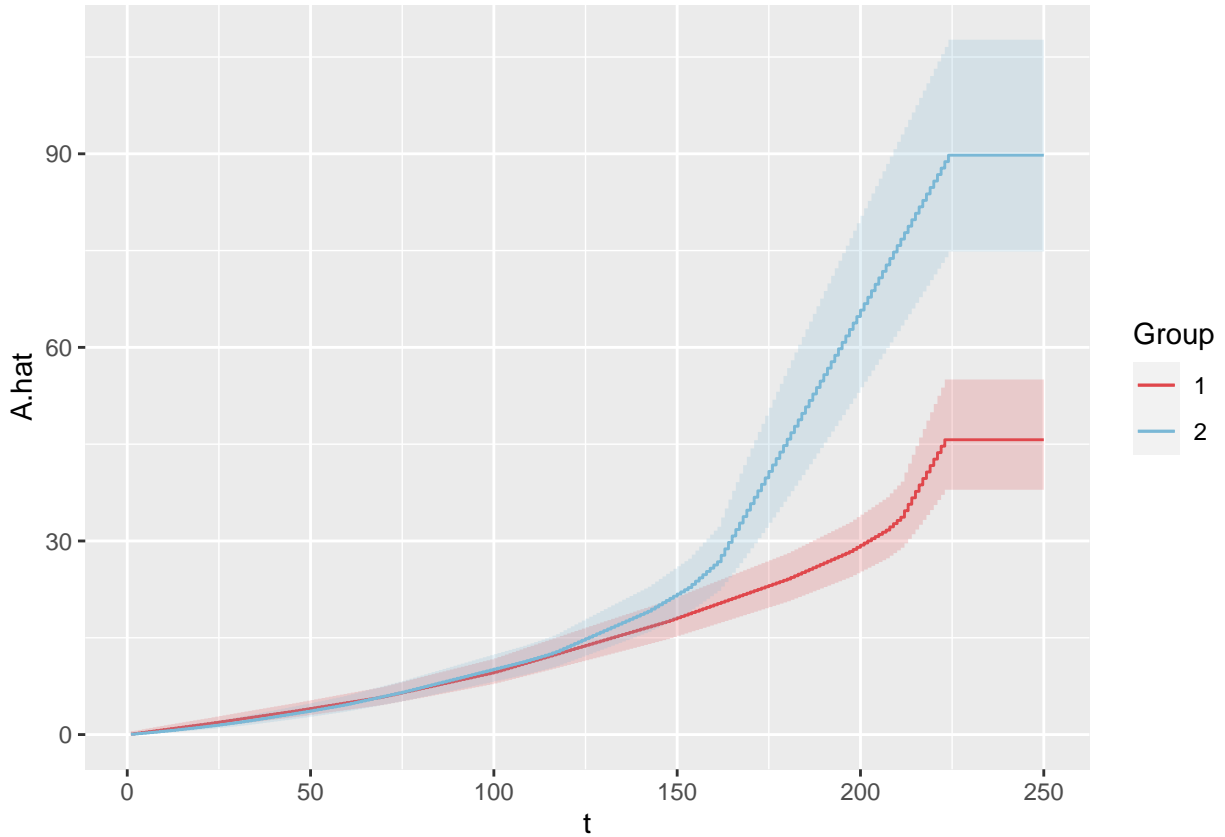


Figure 3: The Nelson-Aalen estimator of the integrated hazard rate with confidence intervals based on the log-transform.

Problem 2