MTH6157 Survival Models

Tutorial Question Week 3 - Solution

1. A researcher is investigating the quality of care for a certain illness in two different hospitals. They collect data from two hospitals on the number of days ten patients from each hospital have to stay before they go home. The results are shown in the table below with * indicating patients who discharge themselves before they recover from the illness.

Hospital A	4	5*	4	4	3	6	5	2	5*	4
Hospital B	2	5	3	2	3	4	7	3	3	7

(a) Use R and the `survival` package to find a Kaplan Meier Estimate of the survival function for each of the hospitals.

first, we create vectors for the hospital stay in days, to indicate censoring and to indicate which hospital.

```
stay <- c(4,5,4,4,3,6,5,4,5,4,2,5,3,2,3,4,7,3,3,7)
leave <- c(1,0,1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,1,1,1)
hospital <- c(0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1)
```

where stay is length in days, leave is 0 for censored data and 1 otherwise, hospital is set to 0 for Hospital A and to 1 for Hospital B.

The two Kaplan Meier estimates for Hospitals A and B can now be found

```
library("survival")
S <- survfit( Surv(stay,leave) ~ hospital, conf.int=0.95, conf.type="plain")
```

assigns the KM estimates to S

(b) Plot the two survival functions making sure that you add suitable axis labels, title, legend and different colours for the two survival functions.

```
plot(S, xlab = "t, days", ylab = "S(t)", main = "KM Survival estimates Hospital A & B", col = c("blue", "red"))
legend("bottomleft", legend = c("A", "B"), col = c("blue", "red"), pch=2)
```

gives the following plot with the KM estimate for Hospital A in blue and for Hospital B in red

KM Survival estimates Hospital A & B
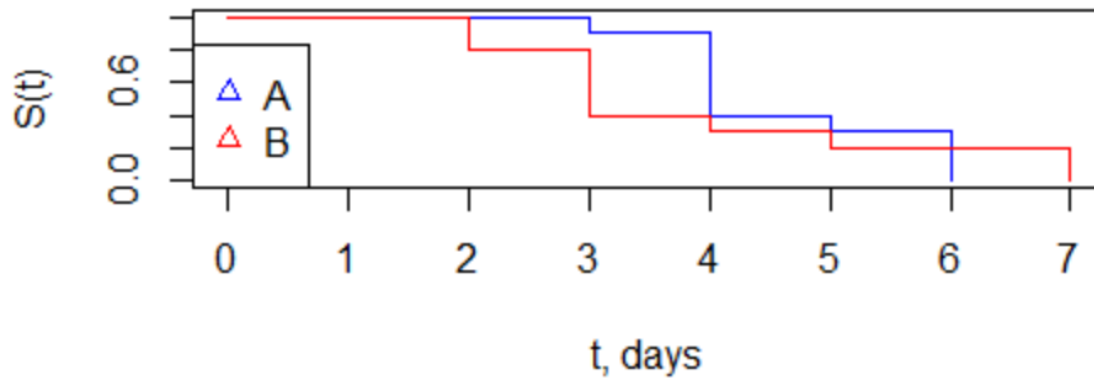
(c) What does this analysis say about which hospital offers the best care for this illness?

We seek to minimise the survival function (shorter hospital stays)

We see from the graph that for $2 \leq t \leq 6$ the survival function for B is lower suggesting that hospital might give better care.

The difference is greatest at 3 and 4 days when B seems to outperform A

However S(t) = 0 at $t \geq 6$ for A but only for $t \geq 7$ at B, so there is a tail of longer stays at B

There is some censoring at A but not at B so that should be investigated