GR5291 - HW9

Mathieu Sauterey - UNI: mjs2364

November 8, 2018

Consider the Mayo Clinic Lung Cancer Data including the variables inst (institution code), time (survival time in days), status (censoring status 1=censored, 2=dead), age (age in years) and sex (male=1 and female=2)

```
# Loads the survival package
library(survival)
```

```
## Warning: package 'survival' was built under R version 3.4.4
```

```
# Loads the Mayo Clinic Lung Cancer Data
data(cancer)
```

```
# Prints the first rows of the dataset
head(cancer)
```

```
##    inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss
## 1     3  306      2  74   1       1       90       100     1175      NA
## 2     3  455      2  68   1       0       90        90     1225      15
## 3     3 1010      1  56   1       0       90        90       NA      15
## 4     5  210      2  57   1       1       90        60     1150      11
## 5     1  883      2  60   1       0      100        90       NA       0
## 6    12 1022      1  74   1       1       50        80      513       0
```
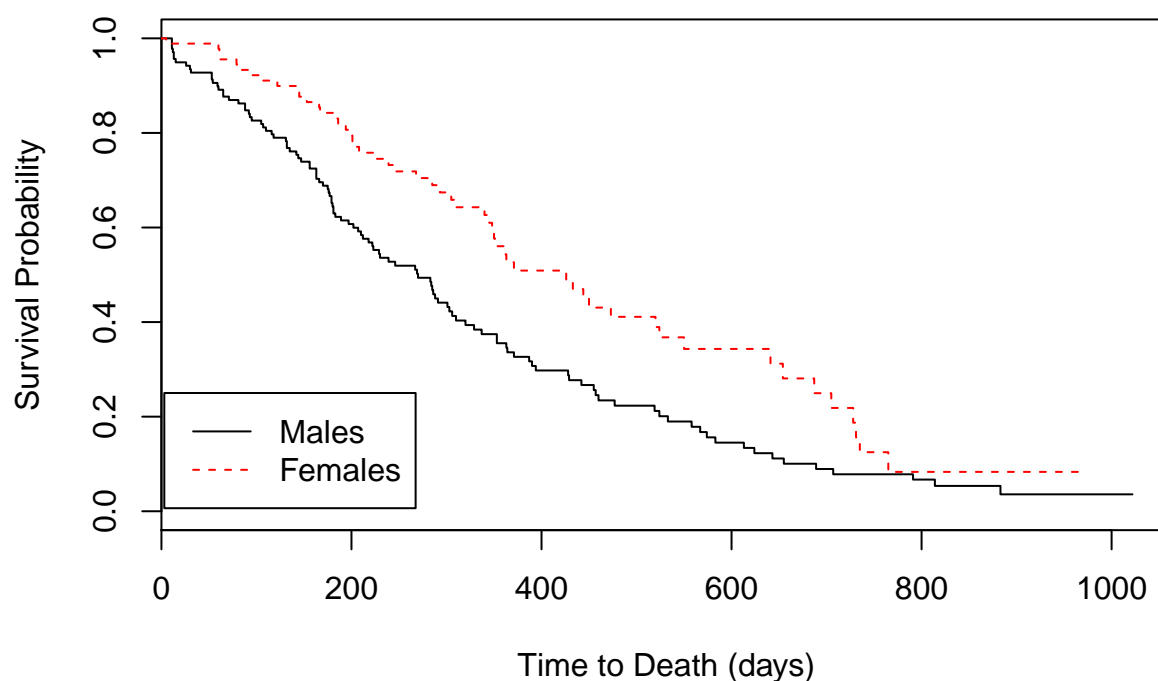
1. Estimate and plot the survival curves for time by sex using:

Kaplan-Meier

```
# Fits the survival curve using the Kaplan-Meier estimate
fit <- survfit(formula = Surv(time, status) ~ sex, data = cancer,
               type = "kaplan-meier")

# Plots the Kaplan-Meier estimated survival curve
plot(fit, lty = 1:2, col = 1:2, xlab = "Time to Death (days)",
     ylab = "Survival Probability",
     main = "Kaplan-Meier estimates for lung cancer data")
legend(x = 3, y = 0.25, legend = c("Males","Females"), lty = 1:2, col = 1:2)
```

Kaplan–Meier estimates for lung cancer data

Survival Probability vs Time to Death (days), with legend: Males (solid black line), Females (red dashed line).

Fleming-Harrington

```r
# Fits the survival curve using the Fleming-Harrington estimate
fit.fh <- survfit(formula = Surv(time, status) ~ sex, data = cancer,
                  type = "fleming-harrington")

# Plots the Fleming-Harrington estimated survival curve
plot(fit.fh, lty = 1:2, col = 1:2, xlab = "Time to Death (days)",
     ylab = "Survival Probability",
     main = "Fleming-Harrington estimates for lung cancer data")
legend(x = 3, y = 0.25, legend = c("Males","Females"), lty = 1:2, col = 1:2)
```

Fleming–Harrington estimates for lung cancer data
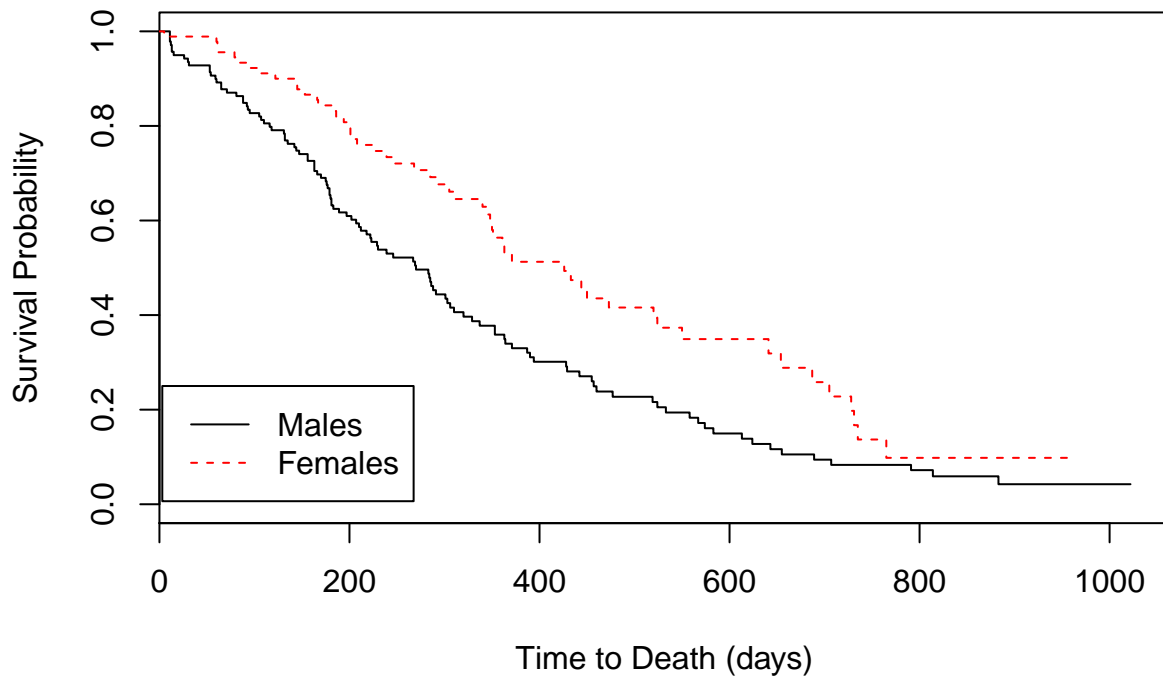
We notice that the survival curves obtained from the Kaplan-Meier and Fleming-Harrington methods are almost identical.

2. For each case in 1, estimate the median survival time, using the estimated survival curves

```
# Median survival time for Kaplan-Meier estimate
print(fit)
```

```
## Call: survfit(formula = Surv(time, status) ~ sex, data = cancer, type = "kaplan-meier")
##
##          n events median 0.95LCL 0.95UCL
## sex=1 138    112    270     212     310
## sex=2  90     53    426     348     550
```

The median survival time obtained from Kaplan-Meier follows from the print above:

Median survival is 270 days for males

Median survival is 426 for females

```
# Median survival time for Fleming-Harrington estimate
fit.fh
```

```
## Call: survfit(formula = Surv(time, status) ~ sex, data = cancer, type = "fleming-harrington")
##
##           n events median 0.95LCL 0.95UCL
```

```
## sex=1 138     112     270     218     320
## sex=2  90      53     426     348     641
```

The median survival time obtained from Fleming-Harrington follows from the print above:

Median survival is 270 days for males

Median survival is 426 for females

We note that the median survival times for both males are females are the same with Kaplan-Meier AND Fleming-Harrington.