5291 hw9

Yijin Wang

11/19/2020

1. Estimate and plot the survival curves for time BY sex using the following methods: · Kaplan-Meier · Fleming-Harrington

```
#install.packages("survival")
library(survival)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

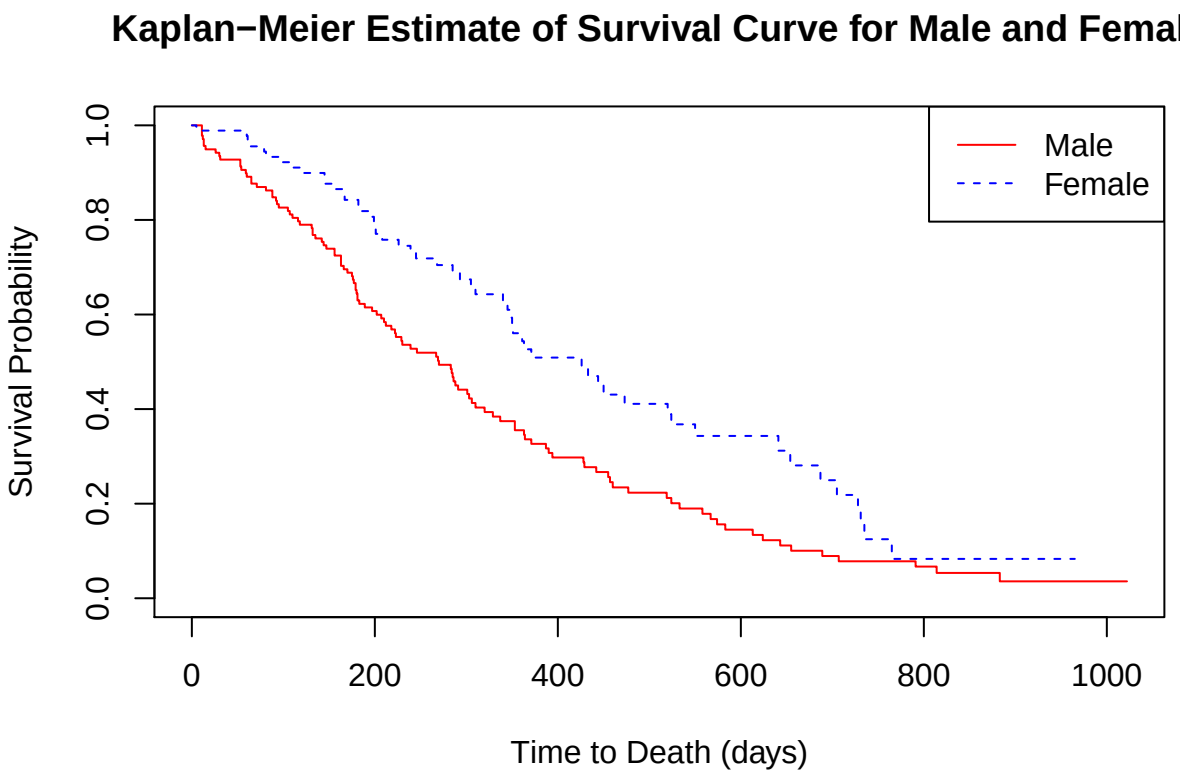
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data(lung)
lung <- lung%>%mutate(sex = as.factor(sex))

#Kaplan-Meier
KM_fit <- survfit(Surv(time, status) ~ sex, data = lung, type = "kaplan-meier" )
KM_fit
```

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

```
plot(KM_fit, col = c("red", "blue"), lty = c(1,2), xlab = "Time to Death (days)", ylab = "Survival Proba
    main = "Kaplan-Meier Estimate of Survival Curve for Male and Female")
legend("topright", c("Male", "Female"), col = c("red", "blue"), lty = c(1,2))
```

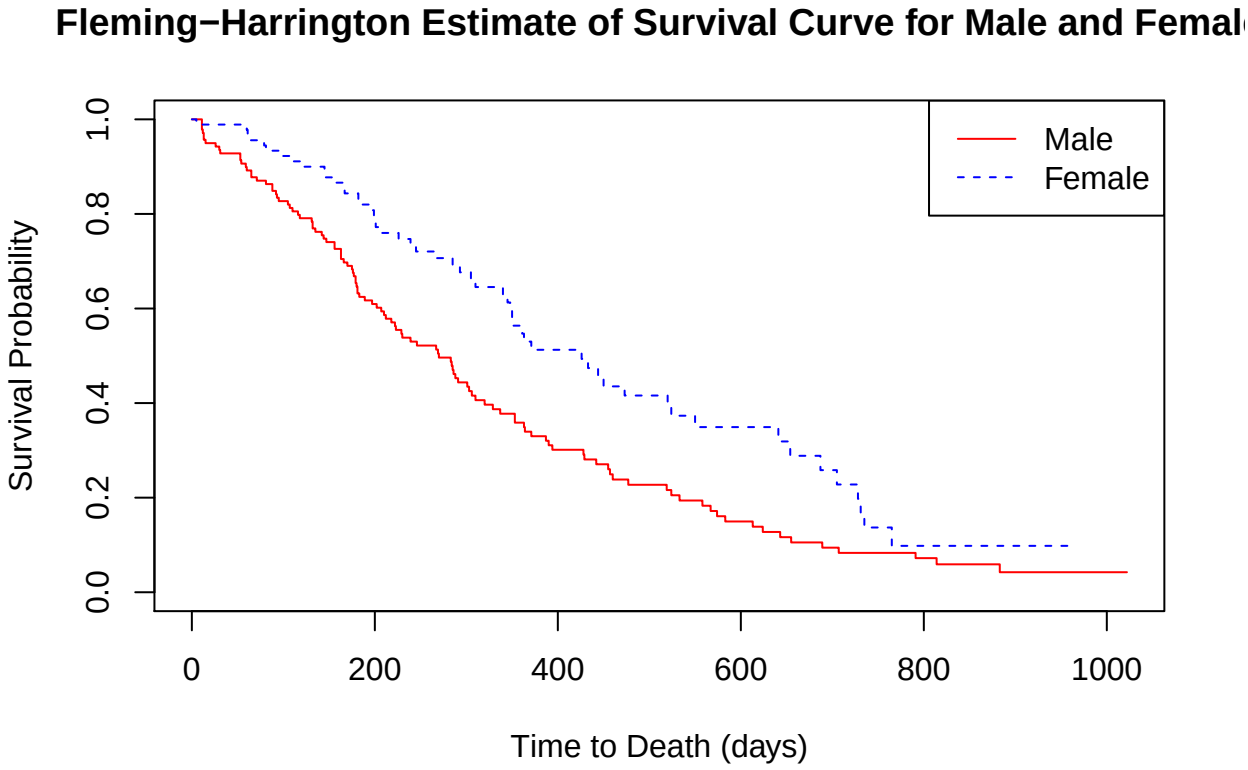


Fleming-Harrington:

```
FH_fit <- survfit(Surv(time, status) ~ sex, data = lung, type = "fleming-harrington")
FH_fit
```

```
## Call: survfit(formula = Surv(time, status) ~ sex, data = lung, type = "fleming-harrington")
##
##         n events median 0.95LCL 0.95UCL
## sex=1 138    112    270     218     320
## sex=2  90     53    426     348     550
```

```
plot(FH_fit, col = c("red", "blue"), lty = c(1,2), xlab = "Time to Death (days)", ylab = "Survival Proba
    main = "Fleming-Harrington Estimate of Survival Curve for Male and Female")
legend("topright", c("Male", "Female"), col = c("red", "blue"), lty = c(1,2))
```



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

```
KM_fit
```

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

```
FH_fit
```

```
## Call: survfit(formula = Surv(time, status) ~ sex, data = lung, type = "fleming-harrington")
##
##         n events median 0.95LCL 0.95UCL
## sex=1 138    112    270     218     320
## sex=2  90     53    426     348     550
```

For both cases, by looking at the survival curves, we conclude that at 0.5 survival probability,the median survival time for male is 270 and the median survival time for female is 426 in both cases.

3. Using a log-rank test, compare the survival distributions for Male and Female

```
logrank <- survdiff(Surv(time, status) ~ sex, data = lung)
logrank
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ sex, data = lung)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 138      112     91.6      4.55      10.3
## sex=2  90       53     73.4      5.68      10.3
##
```

```
##  Chisq= 10.3  on 1 degrees of freedom, p= 0.001
```

Since the p-value = 0.001 < 0.05, we reject the null hypotheses and conclude that there is difference between survival distributions for Male and Female groups.