PSTAT175 Lab B

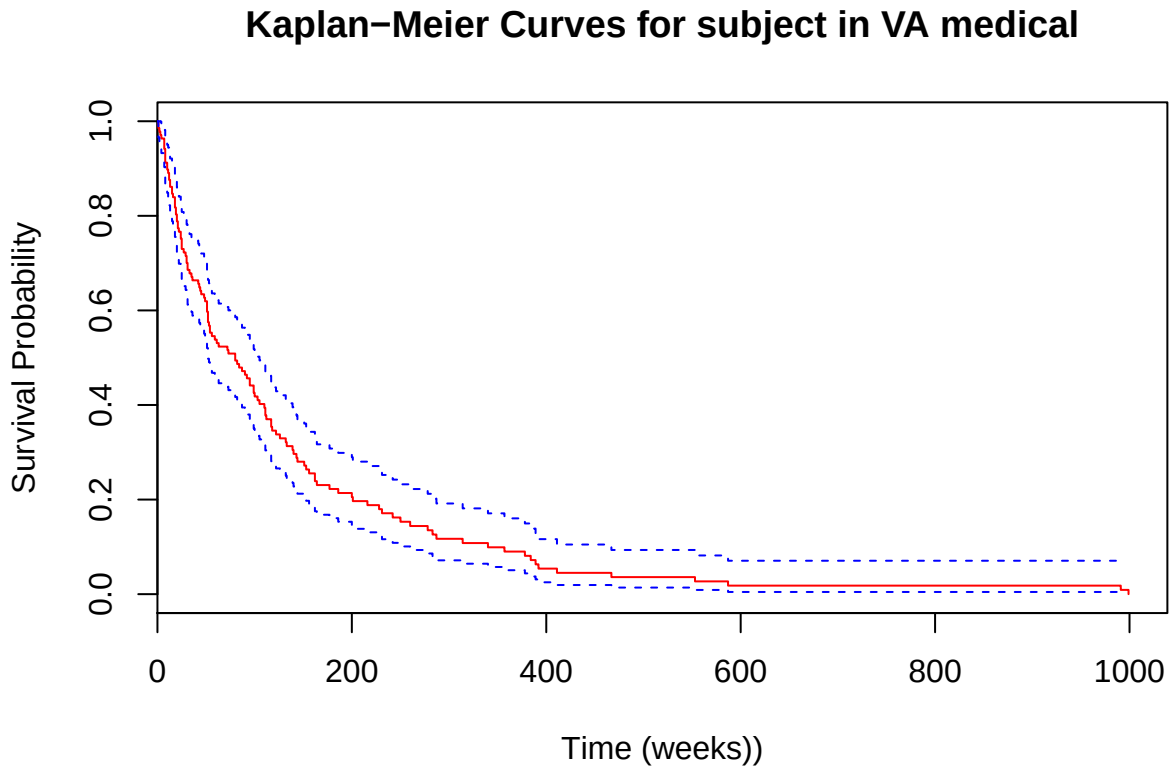
Mujie Wang

2019/10/21

```
#1
#a)
vet <- read.table("/Users/wangmujie/Desktop/vets.txt")
library(survival)
vet.time <- vet$V1
vet.event <- vet$V2
vet.surv <- Surv(vet.time,vet.event)
vet.fit <- survfit(vet.surv ~ 1)
vet.fit
```

```
## Call: survfit(formula = vet.surv ~ 1)
##
##       n  events  median 0.95LCL 0.95UCL
##     137     128      80      52     105
```

```
plot(vet.fit, main= "Kaplan-Meier Curves for subject in VA medical",
     xlab = "Time (weeks))",
     ylab = "Survival Probability",
     col = c("red","blue","blue")
     )
```




```
#b)
q1tm <- min(vet.fit$time[vet.fit$surv < 0.75])
q2tm <- min(vet.fit$time[vet.fit$surv < 0.5])
```

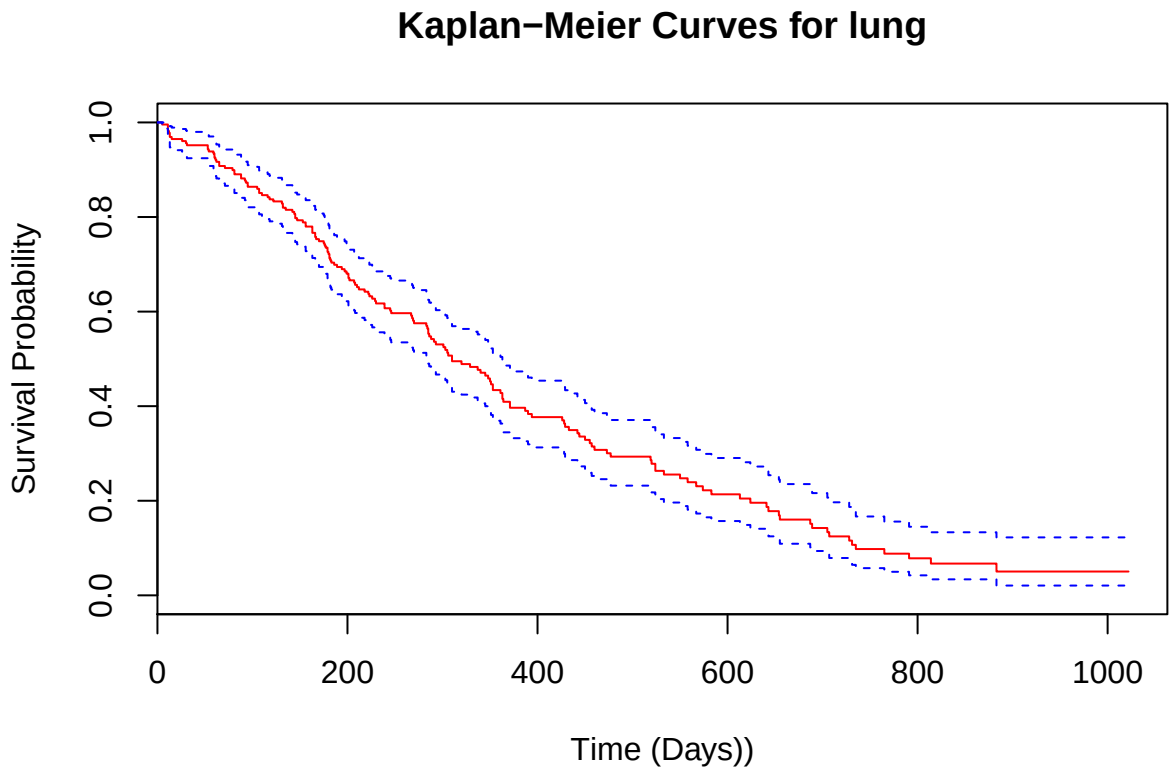
```r
q3tm <- min(vet.fit$time[vet.fit$surv < 0.25])
c(q1tm, q2tm, q3tm)
```

```
## [1]  25  80 162
```

```r
quantile(vet.fit,c(0.25,0.5,0.75))$q
```

```
##  25  50  75
##  25  80 162
```

```r
#2
#a)
library(survival)
data(lung)
lung.fit <- survfit(Surv(lung$time,lung$status) ~ 1)
plot(lung.fit, main= "Kaplan-Meier Curves for lung",
     xlab = "Time (Days))",
     ylab = "Survival Probability",
     conf.int = TRUE,col=c("red","blue","blue")
     )
```




```r
#b)
summary(lung.fit,times=150)
```

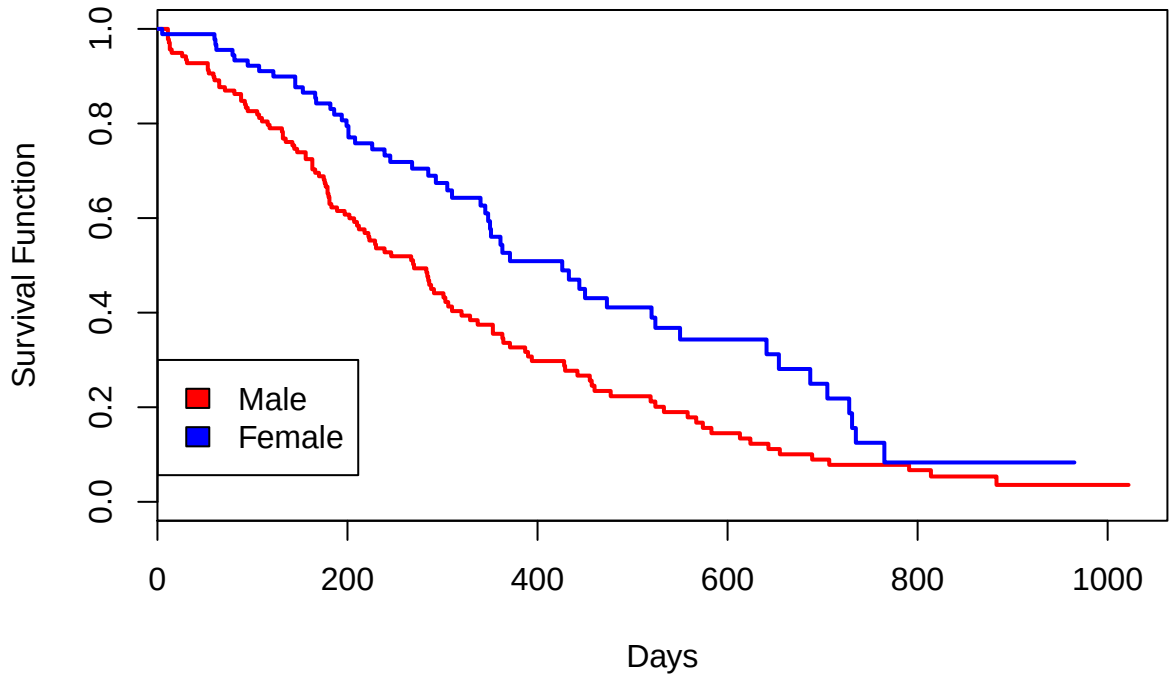
```
## Call: survfit(formula = Surv(lung$time, lung$status) ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   150    179      47    0.793  0.0269        0.742        0.848
```

```r
#According to the data, (0.741,0.848)is the 95% Confidence Interval
#for the survivial function at 150 days
```

```r
#c)
med.tm <- min(lung.fit$time[lung.fit$surv < 0.5])
med.low.tm <- min(lung.fit$time[lung.fit$lower < 0.5])
med.upper.tm <- min(lung.fit$time[lung.fit$upper < 0.5])
c(med.low.tm, med.tm, med.upper.tm)
```

```
## [1] 285 310 363
```

```r
#According to the data, (285,363)is the 95% Confidence Interval for the median survivial time

#d)
sex <- as.factor(lung$sex)
lung.fit.sex <- survfit(Surv(lung$time,lung$status) ~ sex)
plot(lung.fit.sex, col=c(2,4),lwd=2,conf.int= FALSE, xlab = "Days", ylab = "Survival Function")
legend(0,0.3,c("Male","Female"), fill= c(2,4))
```



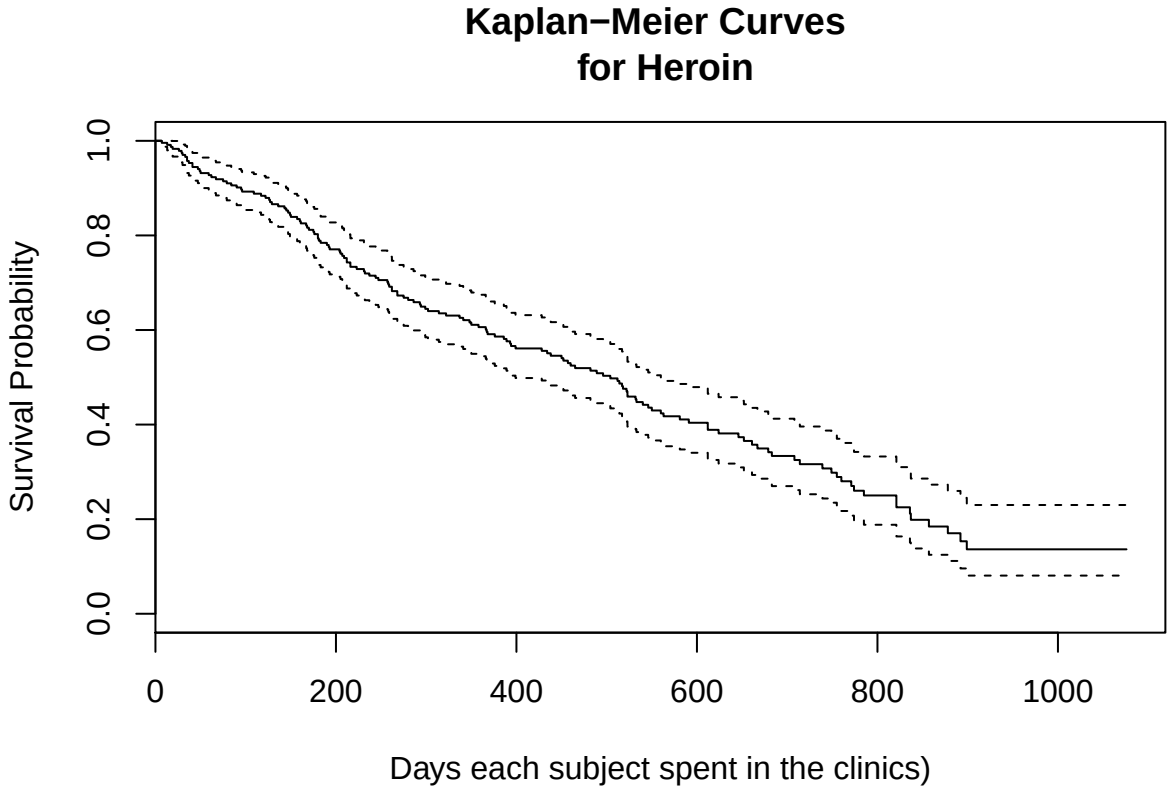

```r
#Generally speaking, women have better suivival rates. By looking through the graph, it is clear that
#the women's curve are above men'a curve.Since a higher survival function means a longer time until
#failure or death, we can say women have better suivival rates.

#e)
quantile(lung.fit.sex,0.5)
```

```
## $quantile
##        50
## sex=1 270
## sex=2 426
##
## $lower
##        50
## sex=1 212
## sex=2 348
##
```

```
## $upper
##        50
## sex=1 310
## sex=2 550
```

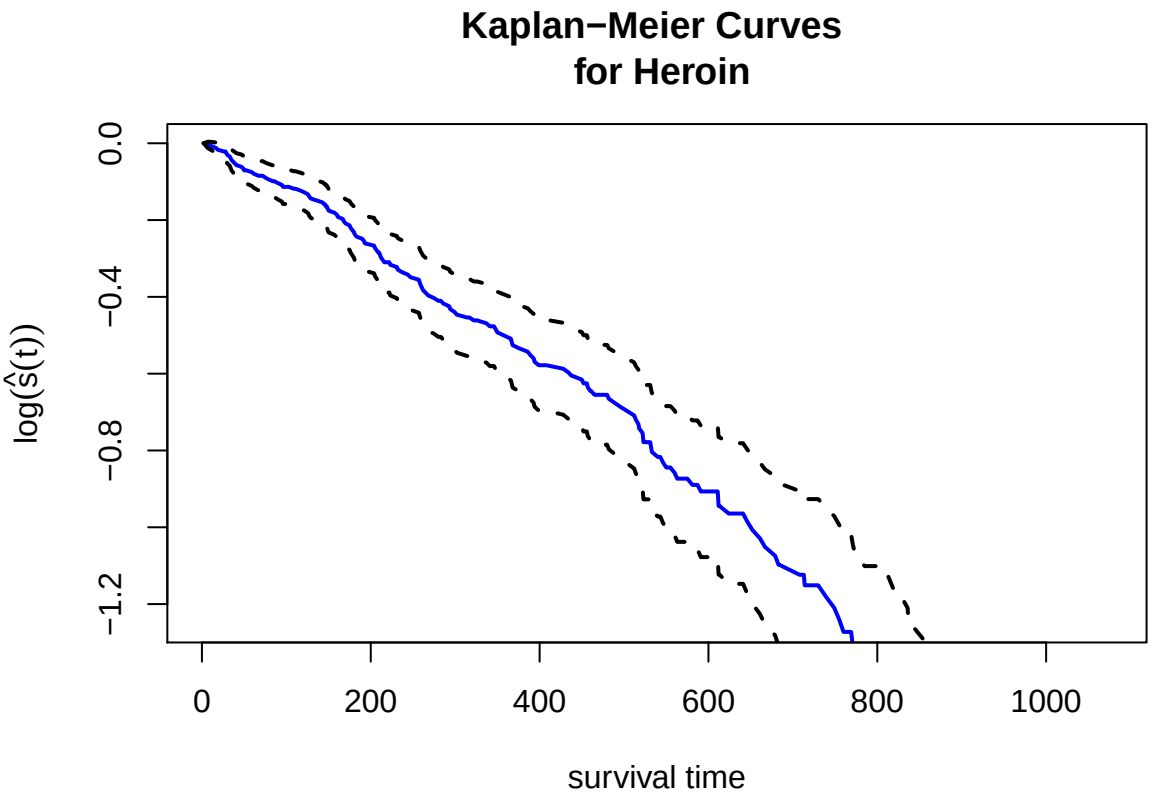
```
#The interval for women is (348,550) while the interval for men is (212,310);
#the median survival time for women is 426 days while for men is 270 days;
#This result agrees with our assesment that women have better suivival rates than men.
#But it may not tell the whole story since the curves come back together at the end which
#indicates the benifit for women may diminish overtime.
```

```
#3
#a)
library(survival)
heroin <- read.table("/Users/wangmujie/desktop/Heroin.txt")
heroin.fit <- survfit(Surv(heroin$Time,heroin$Status) ~ 1)
plot(heroin.fit, main= "Kaplan-Meier Curves \n for Heroin",
     xlab = "Days each subject spent in the clinics)",
     ylab = "Survival Probability",
     conf.int = TRUE
     )
```

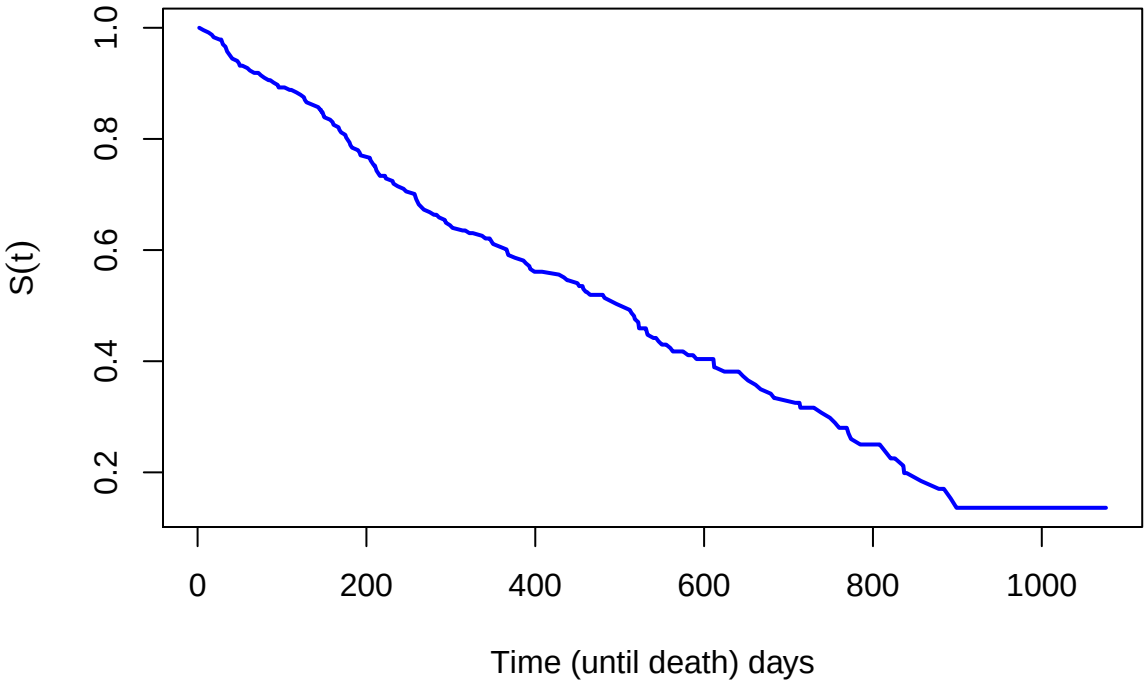



```
#b)
mj = heroin.fit$n.event
nj = heroin.fit$n.risk
Vj=mj/nj/(nj-mj)
cVj=cumsum(Vj)
lowerCI = log(heroin.fit$surv)-1.96*sqrt(cVj)
upperCI = log(heroin.fit$surv)+1.96*sqrt(cVj)
par(mar=c(5,5,4,2))
```

```r
plot(heroin.fit$time,log(heroin.fit$surv),lwd =2, type = "l",
     ylim = c(-1.25,0),
     main= "Kaplan-Meier Curves \n for Heroin",
     xlab = " survival time ",
     ylab = expression(log(hat(s)(t))),col="blue")
lines(heroin.fit$time,lowerCI,lty=2,col=1,lwd=2)
lines(heroin.fit$time,upperCI,lty=2,col=1,lwd=2)
```




```r
#c)
plot(heroin.fit$time,(exp(1))^(log(heroin.fit$surv)),lwd=2, type = "l",
     xlab="Time (until death) days",ylab=expression(S(t)), col=4)
```




```
#d)
summary(heroin.fit,times = 365)
```

```
## Call: survfit(formula = Surv(heroin$Time, heroin$Status) ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   365    122      87    0.606  0.0331        0.545        0.675
```

```
#NULL Hypothesis(H0)is 50% or less of the patients are discharged by the one year mark;
#Estimate is that nearly less than 40% have been discharged at 365 days.
#Therefore, ther is no evidencr for Ha that the survival function is less than 50%

#two-sided test:
heroin.tk <- max(heroin.fit$time[heroin.fit$time <365])
heroin.fit$surv[heroin.fit$time == heroin.tk ]
```

```
## [1] 0.6060647
```

```
zstats <- (log(heroin.fit$surv[heroin.fit$time == heroin.tk]) - log(0.5))/heroin.fit$std.err[heroin.fit
zstats
```

```
## [1] 3.524093
```

```
# zstats > 1.96(critical value while standard error=0.05)
# we can conclude that a statistically significant fewer than 50% are being discharged by the end of th
pnorm(-zstats)
```

```
## [1] 0.0002124677
```

```
#p-value = 0.0002124677

#e)
#95% CI for the 70th percentile
heroin.LI <- min(heroin.fit$time[heroin.fit$lower < 0.3])
heroin.UI <- max(heroin.fit$time[heroin.fit$upper > 0.3])
```

```r
summary(heroin.fit,times = c(heroin.LI,heroin.UI))
```

```
## Call: survfit(formula = Surv(heroin$Time, heroin$Status) ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   661     46     130    0.357  0.0359        0.294        0.435
##   826     18      14    0.225  0.0368        0.163        0.310
```

```r
c(heroin.LI,heroin.UI)
```

```
## [1] 661 826
```

```r
#95% CI for the 70th percentile is (661,826)

#95% CI for the 80th percentile
heroin.LI <- min(heroin.fit$time[heroin.fit$lower < 0.2])
heroin.UI <- max(heroin.fit$time[heroin.fit$upper > 0.2])
summary(heroin.fit,times = c(heroin.LI,1050,heroin.UI))
```

```
## Call: survfit(formula = Surv(heroin$Time, heroin$Status) ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   774     27     141    0.260  0.0364       0.1978        0.342
##  1050      2       9    0.136  0.0364       0.0807        0.230
##  1076      1       0    0.136  0.0364       0.0807        0.230
```

```r
#95% CI for the 80th percentile  is (774,1076)
# tk=1076 is the last time in the whole data set, it is a censored observation.
```