lab2

zhongyun zhang

2019/10/21

Question1

a)Use survfit to plot the Kaplan–Meier estimator of the data along with a 95% confidence interval.

First, we need to import vets.txt Second, we read in the first column of data as a vector and call it vet.time

```
vets <- read.table("~/Desktop/2019 Fall/PSTAT 175/lab2/vets.txt", quote="\"", comment.char="")
vet.time<-vets$V1
```

We read second column as a vector vet.cns, in this column, 1 is an event occured for that patient and 0 is when that patient is censored

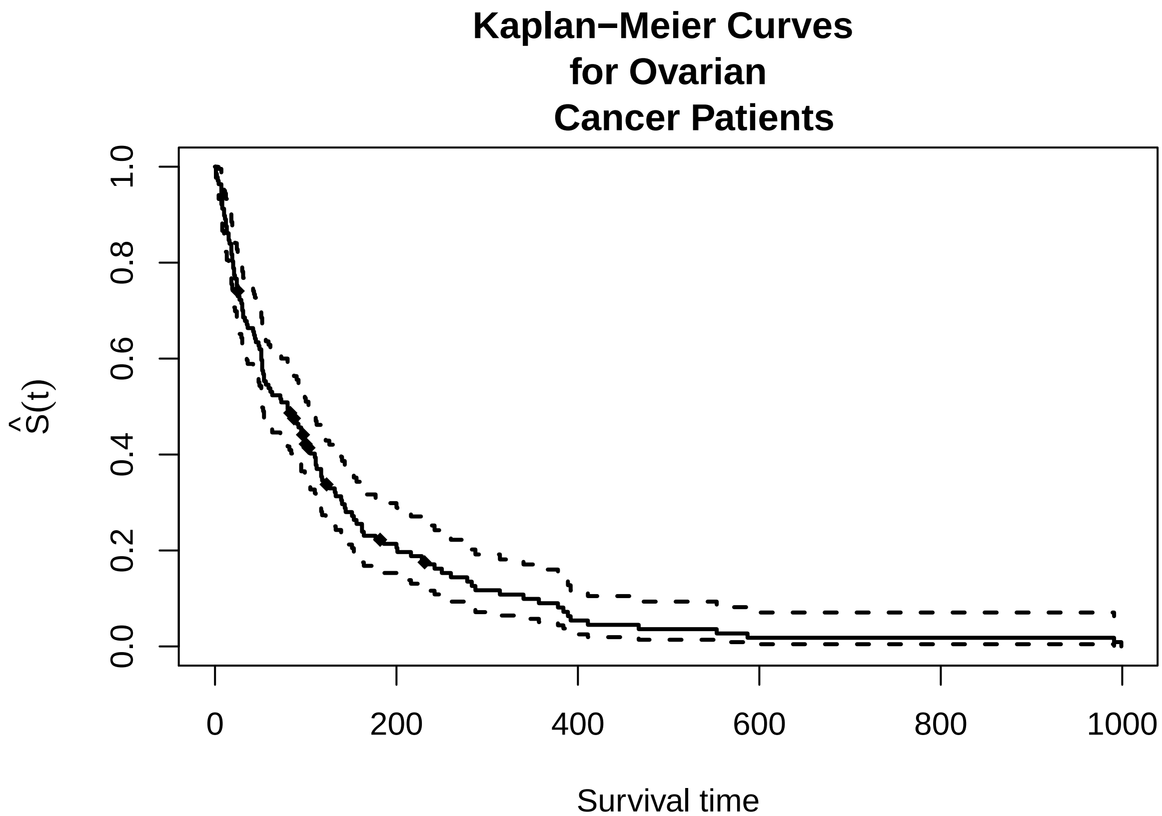
```
vet.cns<-vets$V2
```

we need to load library survival we create a function called vet.surv by using the Surv function.

```
#install.packages("survival")
library(survival)
vet.surv<-Surv(vet.time,vet.cns)
```

Use plot function on the results from survfit to generate a picture of the estimate of the survival function with a 95% confidence interval.

```
vet.fit = survfit(Surv(vet.time,vet.cns)~1,data=vets)
par(mar=c(5,5,4,2))
plot(vet.fit,main="Kaplan-Meier Curves \n for Ovarian
     Cancer Patients",xlab="Survival time",
     ylab=expression(hat(S)(t)),lwd=2, mark.time = TRUE,mark=18)
```



b)Calculate an estimate of the quartiles of this distribution (i.e. the 75th, 50th, and 25th percentiles.)

```
quantile(vet.fit, probs=c(.25,.5,0.75), conf.int=FALSE)
```

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

The estimate of 75th percentile is 162. The estimate of 50th percentile is 80. The estimate of 25th percentile is 25.

Question 2

a)

Construct a survival object ret.surv by using Surv function. First, we need to load library survival

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

Second, we need to import data set lung which is available from the survival library.

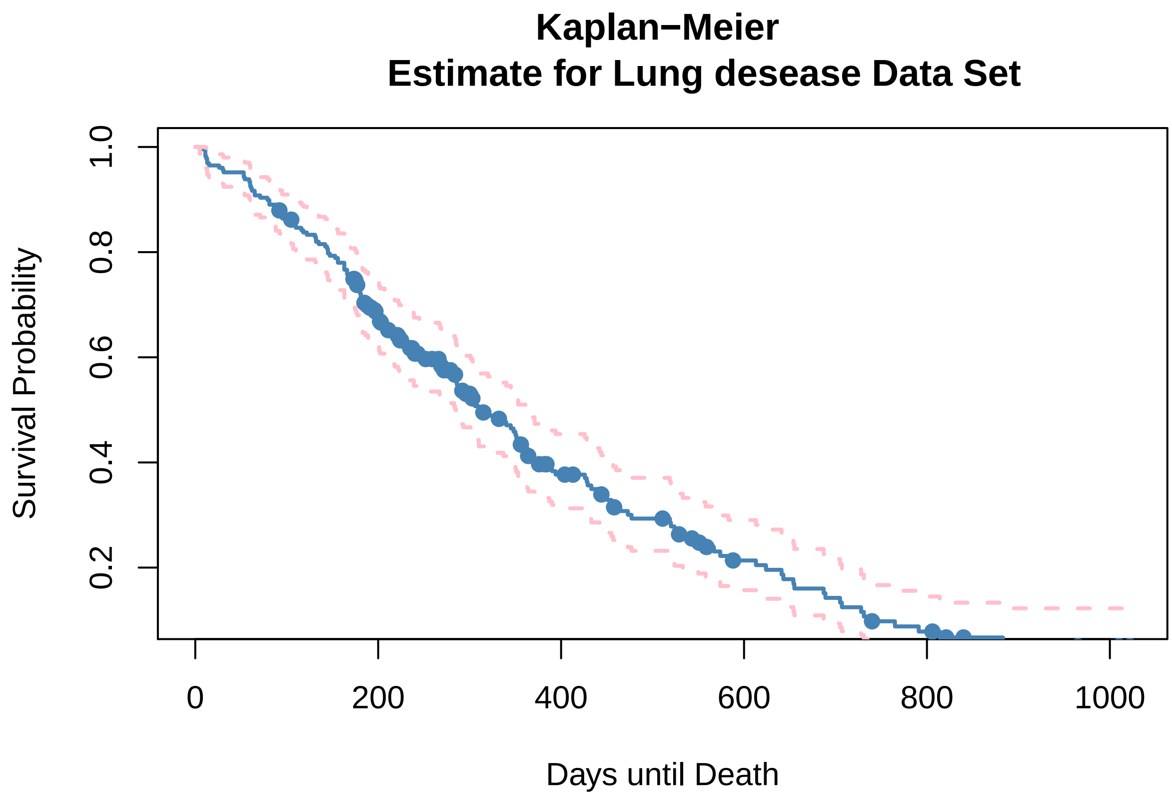
```
data(lung)
```

Then, we can construct surv function

```
lung.surv <- Surv(lung$time,lung$status)
```

Finally, we use survfit to plot the Kaplan–Meier estimator of the data along with a 95% confidence interval.

```r
fit <- survfit(lung.surv ~ 1)
plot(fit,mark=19,lwd=2,ylim = c(0.1,1.0),
        col=c("steelblue","pink","pink"),ylab="Survival Probability",
        xlab="Days until Death", main="Kaplan-Meier
        Estimate for Lung desease Data Set")
```



b) We calculate an estimate and a 95% confidence interval for the survivor function at 150 days.

There is no event happend in day 150. Assume that the survival time at 150 is the mase as the survival time at the previous and nearest time point when the event happened. Then we can get 95% confidence interval for survival function at 150 days.

```r
summary(fit)
```

```
## Call: survfit(formula = lung.surv ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     5    228       1   0.9956 0.00438       0.9871        1.000
##    11    227       3   0.9825 0.00869       0.9656        1.000
##    12    224       1   0.9781 0.00970       0.9592        0.997
##    13    223       2   0.9693 0.01142       0.9472        0.992
##    15    221       1   0.9649 0.01219       0.9413        0.989
##    26    220       1   0.9605 0.01290       0.9356        0.986
##    30    219       1   0.9561 0.01356       0.9299        0.983
##    31    218       1   0.9518 0.01419       0.9243        0.980
##    53    217       2   0.9430 0.01536       0.9134        0.974
##    54    215       1   0.9386 0.01590       0.9079        0.970
##    59    214       1   0.9342 0.01642       0.9026        0.967
```

```
##    60    213       2   0.9254 0.01740       0.8920        0.960
##    61    211       1   0.9211 0.01786       0.8867        0.957
##    62    210       1   0.9167 0.01830       0.8815        0.953
##    65    209       2   0.9079 0.01915       0.8711        0.946
##    71    207       1   0.9035 0.01955       0.8660        0.943
##    79    206       1   0.8991 0.01995       0.8609        0.939
##    81    205       2   0.8904 0.02069       0.8507        0.932
##    88    203       2   0.8816 0.02140       0.8406        0.925
##    92    201       1   0.8772 0.02174       0.8356        0.921
##    93    199       1   0.8728 0.02207       0.8306        0.917
##    95    198       2   0.8640 0.02271       0.8206        0.910
##   105    196       1   0.8596 0.02302       0.8156        0.906
##   107    194       2   0.8507 0.02362       0.8056        0.898
##   110    192       1   0.8463 0.02391       0.8007        0.894
##   116    191       1   0.8418 0.02419       0.7957        0.891
##   118    190       1   0.8374 0.02446       0.7908        0.887
##   122    189       1   0.8330 0.02473       0.7859        0.883
##   131    188       1   0.8285 0.02500       0.7810        0.879
##   132    187       2   0.8197 0.02550       0.7712        0.871
##   135    185       1   0.8153 0.02575       0.7663        0.867
##   142    184       1   0.8108 0.02598       0.7615        0.863
##   144    183       1   0.8064 0.02622       0.7566        0.859
##   145    182       2   0.7975 0.02667       0.7469        0.852
##   147    180       1   0.7931 0.02688       0.7421        0.848
##   153    179       1   0.7887 0.02710       0.7373        0.844
##   156    178       2   0.7798 0.02751       0.7277        0.836
##   163    176       3   0.7665 0.02809       0.7134        0.824
##   166    173       2   0.7577 0.02845       0.7039        0.816
##   167    171       1   0.7532 0.02863       0.6991        0.811
##   170    170       1   0.7488 0.02880       0.6944        0.807
##   175    167       1   0.7443 0.02898       0.6896        0.803
##   176    165       1   0.7398 0.02915       0.6848        0.799
##   177    164       1   0.7353 0.02932       0.6800        0.795
##   179    162       2   0.7262 0.02965       0.6704        0.787
##   180    160       1   0.7217 0.02981       0.6655        0.783
##   181    159       2   0.7126 0.03012       0.6559        0.774
##   182    157       1   0.7081 0.03027       0.6511        0.770
##   183    156       1   0.7035 0.03041       0.6464        0.766
##   186    154       1   0.6989 0.03056       0.6416        0.761
##   189    152       1   0.6943 0.03070       0.6367        0.757
##   194    149       1   0.6897 0.03085       0.6318        0.753
##   197    147       1   0.6850 0.03099       0.6269        0.749
##   199    145       1   0.6803 0.03113       0.6219        0.744
##   201    144       2   0.6708 0.03141       0.6120        0.735
##   202    142       1   0.6661 0.03154       0.6071        0.731
##   207    139       1   0.6613 0.03168       0.6020        0.726
##   208    138       1   0.6565 0.03181       0.5970        0.722
##   210    137       1   0.6517 0.03194       0.5920        0.717
##   212    135       1   0.6469 0.03206       0.5870        0.713
##   218    134       1   0.6421 0.03218       0.5820        0.708
##   222    132       1   0.6372 0.03231       0.5769        0.704
##   223    130       1   0.6323 0.03243       0.5718        0.699
##   226    126       1   0.6273 0.03256       0.5666        0.694
##   229    125       1   0.6223 0.03268       0.5614        0.690
```

```
##   230    124       1   0.6172 0.03280       0.5562        0.685
##   239    121       2   0.6070 0.03304       0.5456        0.675
##   245    117       1   0.6019 0.03316       0.5402        0.670
##   246    116       1   0.5967 0.03328       0.5349        0.666
##   267    112       1   0.5913 0.03341       0.5294        0.661
##   268    111       1   0.5860 0.03353       0.5239        0.656
##   269    110       1   0.5807 0.03364       0.5184        0.651
##   270    108       1   0.5753 0.03376       0.5128        0.645
##   283    104       1   0.5698 0.03388       0.5071        0.640
##   284    103       1   0.5642 0.03400       0.5014        0.635
##   285    101       2   0.5531 0.03424       0.4899        0.624
##   286     99       1   0.5475 0.03434       0.4841        0.619
##   288     98       1   0.5419 0.03444       0.4784        0.614
##   291     97       1   0.5363 0.03454       0.4727        0.608
##   293     94       1   0.5306 0.03464       0.4669        0.603
##   301     91       1   0.5248 0.03475       0.4609        0.597
##   303     89       1   0.5189 0.03485       0.4549        0.592
##   305     87       1   0.5129 0.03496       0.4488        0.586
##   306     86       1   0.5070 0.03506       0.4427        0.581
##   310     85       2   0.4950 0.03523       0.4306        0.569
##   320     82       1   0.4890 0.03532       0.4244        0.563
##   329     81       1   0.4830 0.03539       0.4183        0.558
##   337     79       1   0.4768 0.03547       0.4121        0.552
##   340     78       1   0.4707 0.03554       0.4060        0.546
##   345     77       1   0.4646 0.03560       0.3998        0.540
##   348     76       1   0.4585 0.03565       0.3937        0.534
##   350     75       1   0.4524 0.03569       0.3876        0.528
##   351     74       1   0.4463 0.03573       0.3815        0.522
##   353     73       2   0.4340 0.03578       0.3693        0.510
##   361     70       1   0.4278 0.03581       0.3631        0.504
##   363     69       2   0.4154 0.03583       0.3508        0.492
##   364     67       1   0.4092 0.03582       0.3447        0.486
##   371     65       2   0.3966 0.03581       0.3323        0.473
##   387     60       1   0.3900 0.03582       0.3258        0.467
##   390     59       1   0.3834 0.03582       0.3193        0.460
##   394     58       1   0.3768 0.03580       0.3128        0.454
##   426     55       1   0.3700 0.03580       0.3060        0.447
##   428     54       1   0.3631 0.03579       0.2993        0.440
##   429     53       1   0.3563 0.03576       0.2926        0.434
##   433     52       1   0.3494 0.03573       0.2860        0.427
##   442     51       1   0.3426 0.03568       0.2793        0.420
##   444     50       1   0.3357 0.03561       0.2727        0.413
##   450     48       1   0.3287 0.03555       0.2659        0.406
##   455     47       1   0.3217 0.03548       0.2592        0.399
##   457     46       1   0.3147 0.03539       0.2525        0.392
##   460     44       1   0.3076 0.03530       0.2456        0.385
##   473     43       1   0.3004 0.03520       0.2388        0.378
##   477     42       1   0.2933 0.03508       0.2320        0.371
##   519     39       1   0.2857 0.03498       0.2248        0.363
##   520     38       1   0.2782 0.03485       0.2177        0.356
##   524     37       2   0.2632 0.03455       0.2035        0.340
##   533     34       1   0.2554 0.03439       0.1962        0.333
##   550     32       1   0.2475 0.03423       0.1887        0.325
##   558     30       1   0.2392 0.03407       0.1810        0.316
```

```
##   567      28        1   0.2307 0.03391        0.1729        0.308
##   574      27        1   0.2221 0.03371        0.1650        0.299
##   583      26        1   0.2136 0.03348        0.1571        0.290
##   613      24        1   0.2047 0.03325        0.1489        0.281
##   624      23        1   0.1958 0.03297        0.1407        0.272
##   641      22        1   0.1869 0.03265        0.1327        0.263
##   643      21        1   0.1780 0.03229        0.1247        0.254
##   654      20        1   0.1691 0.03188        0.1169        0.245
##   655      19        1   0.1602 0.03142        0.1091        0.235
##   687      18        1   0.1513 0.03090        0.1014        0.226
##   689      17        1   0.1424 0.03034        0.0938        0.216
##   705      16        1   0.1335 0.02972        0.0863        0.207
##   707      15        1   0.1246 0.02904        0.0789        0.197
##   728      14        1   0.1157 0.02830        0.0716        0.187
##   731      13        1   0.1068 0.02749        0.0645        0.177
##   735      12        1   0.0979 0.02660        0.0575        0.167
##   765      10        1   0.0881 0.02568        0.0498        0.156
##   791       9        1   0.0783 0.02462        0.0423        0.145
##   814       7        1   0.0671 0.02351        0.0338        0.133
##   883       4        1   0.0503 0.02285        0.0207        0.123
```

```r
max(lung$time[lung$time<150])
```

```
## [1] 147
```

```r
summary(fit,time=150)
```

```
## Call: survfit(formula = lung.surv ~ 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
```

We can say that we are 95% confident that the survival rate at 150th day is between 0.742 and 0.848.

c) We calculate an estimate and a 95% confidence interval for the median survival time.

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

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

```r
print(fit)
```

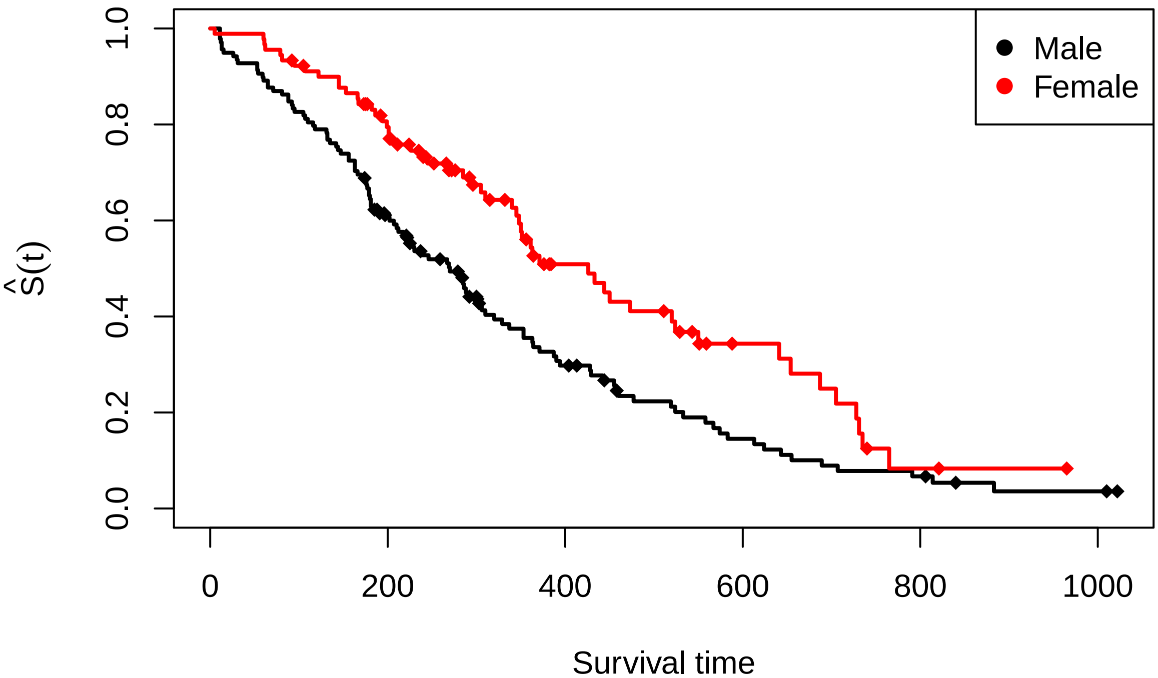
```
## Call: survfit(formula = lung.surv ~ 1)
##
##        n  events  median 0.95LCL 0.95UCL
##      228     165     310     285     363
```

The estimate of median suvival time is 310. The 95% confidence interval is (285,363). We are 95% confidnet that the median survival time is between 285 and 363.

d)

Plot separate estimators of the survival function for men and women.

```r
sexmf = survfit(lung.surv~sex,data=lung)
par(mar=c(5,5,4,2))
plot(sexmf,xlab="Survival time",ylab = expression(hat(S)(t)),lwd=2, col=1:2, mark.time = TRUE,mark=18)
legend("topright",legend=c("Male","Female"),col=1:2,pch=rep(19,2))
```



Generally speaking, the estimate for the women' survial rate is higher. A higher survival function means a longer time until failure or death. There is a small area right at the beginning where function for women is less than the one for men, but this probably represents only a couple of failures. The curves converge near the end of the survival time, around days 750, which means that men and women may have nearly similar survival rates.

e)

Calculate separate estimates and 95% confidence intervals for the median survival time for men and women.

```r
#or we can also use survfit(Surv(lung$time,lung$status)~1, data=lung)
quantile(sexmf,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 estimate for median survival time for men is 270, the 95% interval is (212,310). We are 95% confident that the median survival time for men is between 212 and 310. The estimate for median survival time for wemen is 426, the 95% interval is (348,550). We are 95% confident that the median survival time for women is between 310 and 550. We can see that the lower bound confident interval for women is even hight than the higher confident interval for man. It shows that, generally speaking, the survival rate for female is higher thant he survival rate for men. However, it cannot represent the whole stroy because this is only the 95% confidence intervals for the median survival time for men and women. We cannot use it to assume survival rate for other percentile.

Question 3

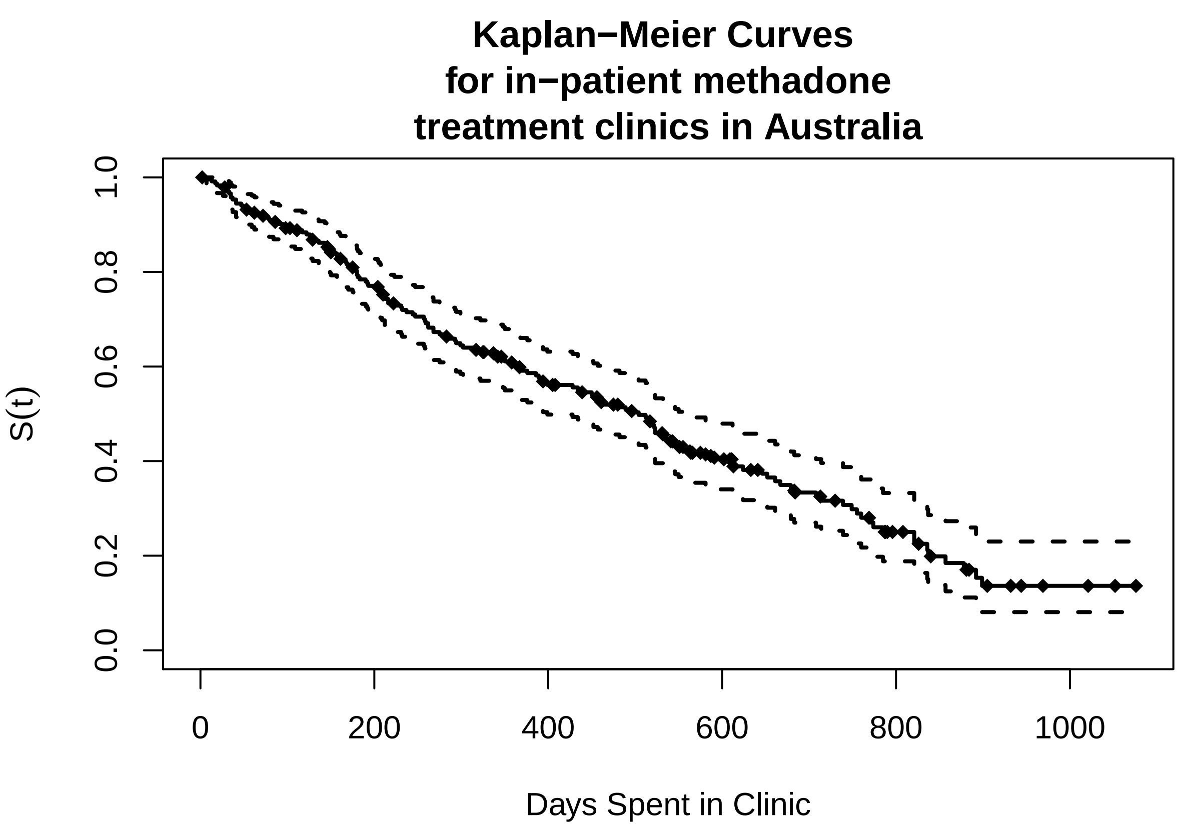
a)

First, we read in the data

```
Heroin <- read.table("~/Desktop/2019 Fall/PSTAT 175/lab2/Heroin.txt", quote="\"", comment.char="")
```

then survfit to plot the Kaplan–Meier estimate along with its 95% confidence interval.

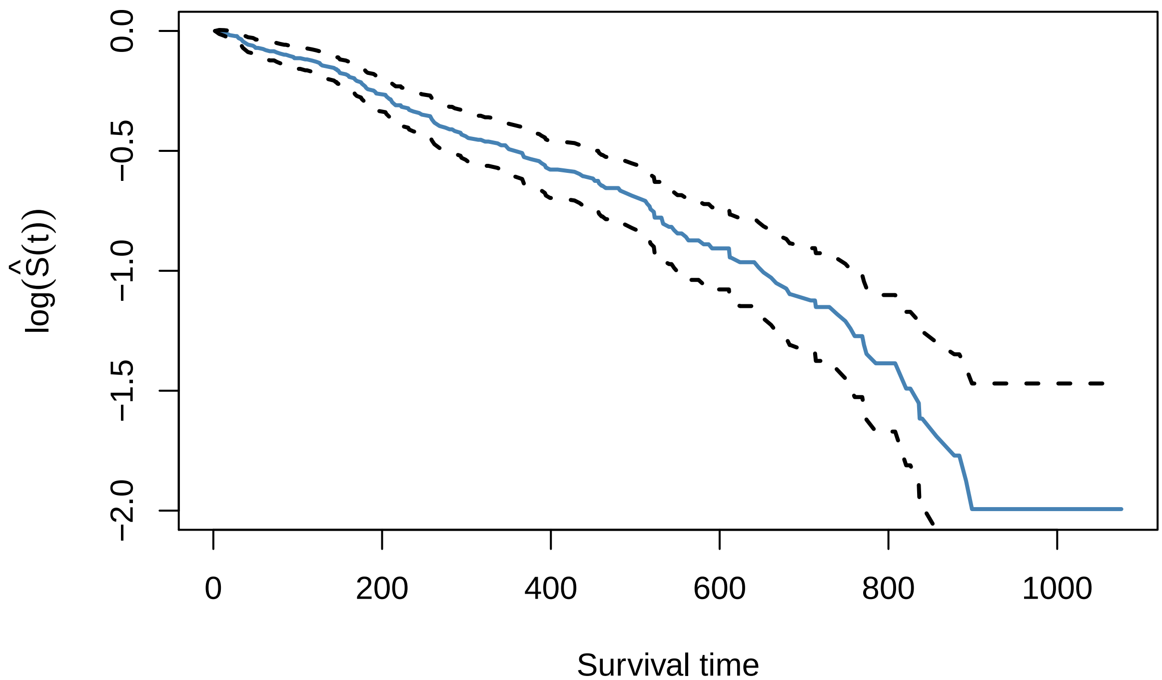
```
heroin.surv <- Surv(Heroin$Time,Heroin$Status)
heroin.fit <- survfit(heroin.surv ~ 1)
plot(heroin.fit,main="Kaplan-Meier Curves \n for in-patient methadone
treatment clinics in Australia",xlab="Days Spent in Clinic",ylab=expression(hat(S)(t)),lwd=2, mark.time
```



b)Plot an estimate of log(S^(t)) along with a 95% confidence interval computed directly from Greenwood's formula for the variance.

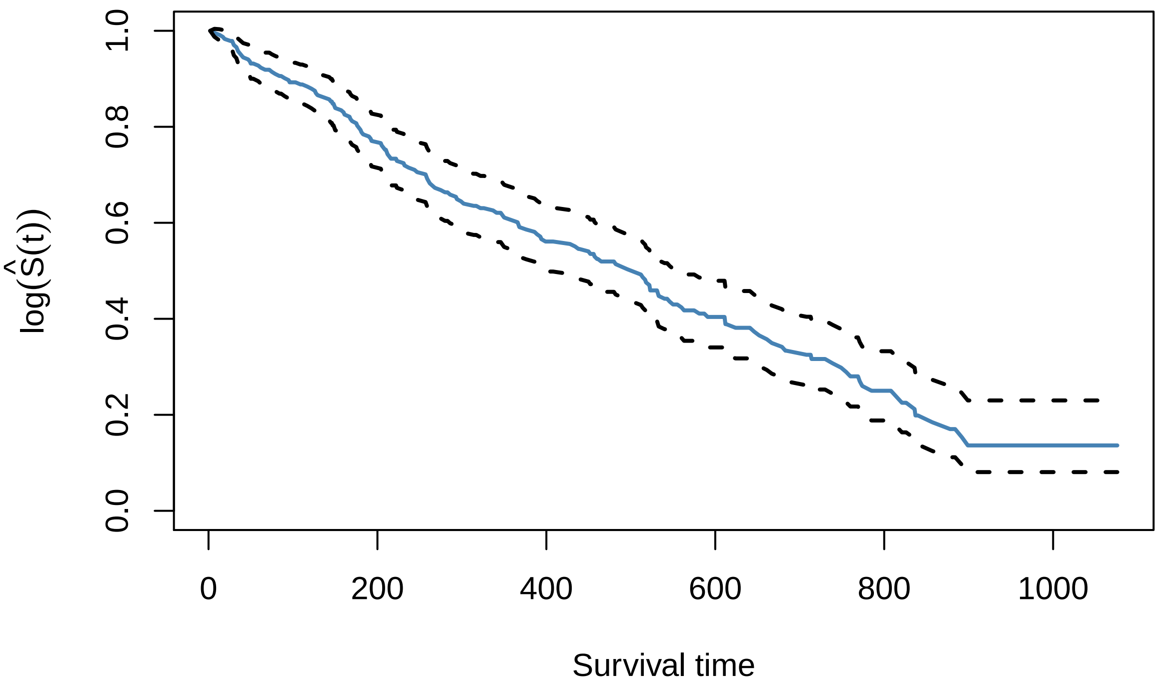
```
mj = heroin.fit$n.event
nj = heroin.fit$n.risk
```

```
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))
plot(heroin.fit$time,log(heroin.fit$surv),lwd=2,type="l",ylim=c(-2,0),xlab="Survival time",ylab=express
lines(heroin.fit$time,lowerCI,lty=2,col=1,lwd=2)
lines(heroin.fit$time,upperCI,lty=2,col=1,lwd=2)
```



c)Transform the plot from part b to give a plot of the estimate of S(t) with a 95% confidence interval. Compare this picture to the result from part a.

```
par(mar=c(5,5,4,2))
plot(heroin.fit$time,exp(log(heroin.fit$surv)),lwd=2,type="l",
     ylim=c(0,1),xlab="Survival time",ylab=expression(log(hat(S)(t))),col="steelblue")
lines(heroin.fit$time,exp(lowerCI),lty=2,col=1,lwd=2)
lines(heroin.fit$time,exp(upperCI),lty=2,col=1,lwd=2)
```



When we compare the picture to the result from part a, we can see that these two pictures are quite similar.

d)

```r
summary(heroin.fit, times=365)
```

```
## Call: survfit(formula = heroin.surv ~ 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
```

```r
twoside.h <- max(heroin.fit$time[heroin.fit$time < 365])
sample_s=0.606
se=0.0331
assumed_s=0.5
TS<-(sample_s-assumed_s)/se
TS
```

```
## [1] 3.202417
```

```r
p_value=pnorm(TS,lower.tail=T)
p_value
```

```
## [1] 0.9993186
```

H0: S(365)>=0.5 (At least 50% of the patients are in the clinic for more than one year) Ha: S(365)<0.5 (less than 50% of the patients are in the clinis for more than one year) We use the one-sided test since we only need the p-value for P(Z>z). Use a significant level of 0.05, since the p value=0.9993186, which is much greater than 0.95, we fail to reject the null hypothesis. We can conclude that at least 50% of the patients are in the clinic for more than one year.

e)

Using the 95% confidence intervals produced by the summary of the survfit output, give an interval for the 70th percentile.

```r
lower.h <- min(heroin.fit$time[heroin.fit$lower < 0.3])
upper.h <- max(heroin.fit$time[heroin.fit$upper > 0.3])
summary(heroin.fit,time=c(lower.h,upper.h))
```

```
## Call: survfit(formula = heroin.surv ~ 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(lower.h,upper.h)
```

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

The interval for 70th percentile along with 95% confident interval is (661,826)

```r
quantile(heroin.fit,probs=0.8,conf.int = TRUE)
```

```
## $quantile
##  80
## 837
##
## $lower
##  80
## 774
##
## $upper
## 80
## NA
```

The difficulty I find if I tried to find an interval for the 80th percentile is that it gives a missing value for the upper bound.