

PSTAT 175 Lab C

Yanjie Qi

2019/10/27

1

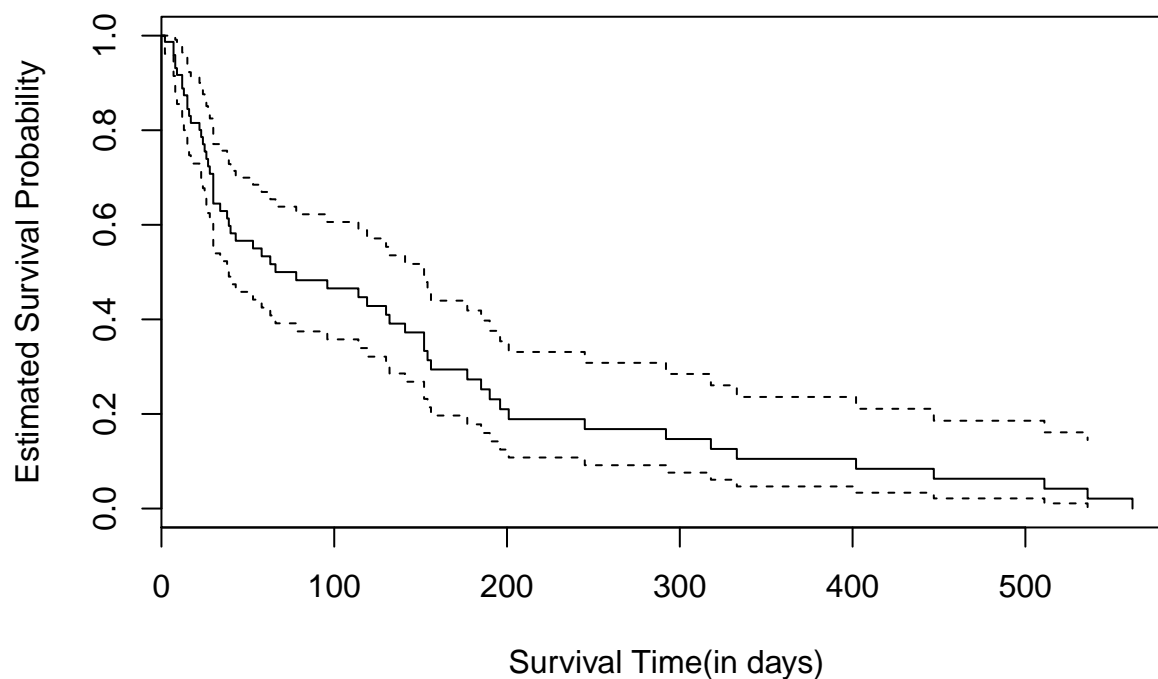
Load the required library and dataset:

```
library(survival)
data(kidney)
attach(kidney)
```

(a)

Plot the KM estimate:

```
## vector
kid.time <- kidney$time
kid.cns <- kidney$status
kid.surv <- Surv(kid.time, kid.cns)
kid.fit <- survfit(kid.surv ~ 1)
## plot
plot(kid.fit, xlab="Survival Time(in days)",
     ylab="Estimated Survival Probability")
```



(b)

Perform Logrank test:

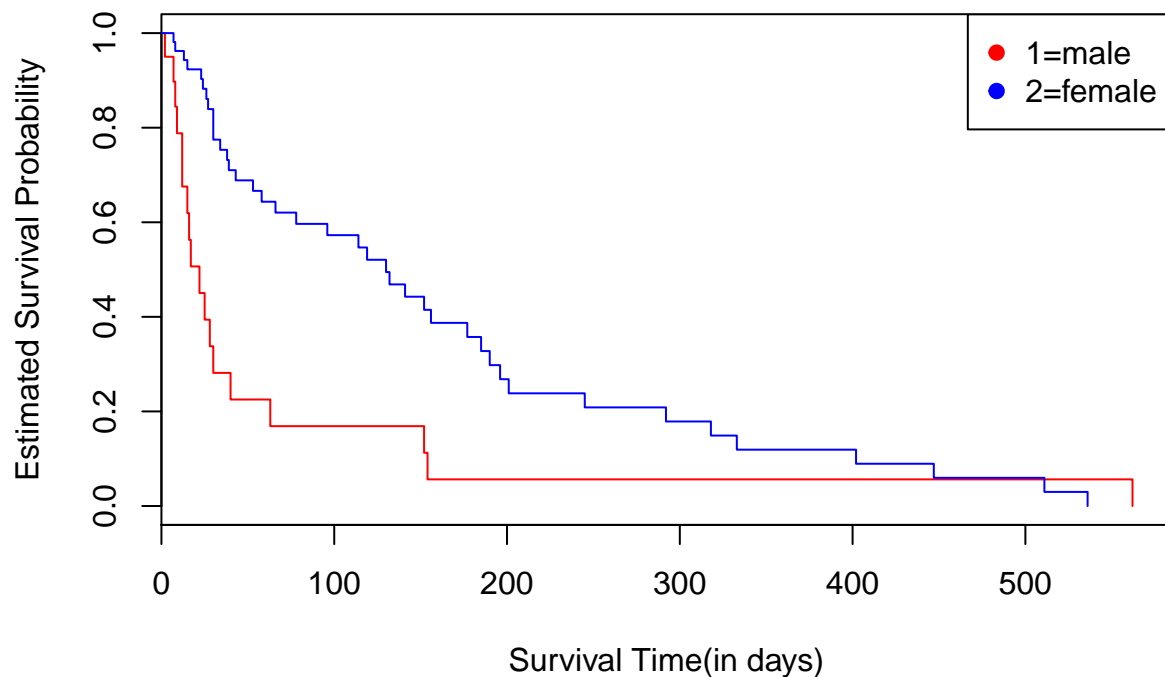
```
survdif(kid.surv ~ sex)
```

```
## Call:
## survdiff(formula = kid.surv ~ sex)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 20      18      10.2      5.99      8.31
## sex=2 56      40      47.8      1.28      8.31
##
##  Chisq= 8.3  on 1 degrees of freedom, p= 0.004
```

The p-value is 0.004, which means that there is significant difference between male and female.

(c)

```
kid.fits <- survfit(kid.surv ~ sex, data = kidney)
plot(kid.fits,xlab="Survival Time(in days)",
     ylab="Estimated Survival Probability", col=c("red","blue"))
legend("topright",legend=c("1=male","2=female"),
     col=c("red","blue"),pch=rep(19,2))
```



As shown by the plot above, there is obvious gap between male line and female line especially between time 10 to 350 so that we could confirm the significant difference between male and female.

(d)

Use coxph to compare male and female group:

```
cox0 <- coxph(kid.surv ~ sex, data = kidney)
cox0
```

```
## Call:
## coxph(formula = kid.surv ~ sex, data = kidney)
##
##           coef exp(coef) se(coef)      z      p
## sex -0.8377    0.4327    0.2966 -2.824 0.00474
```

```
##
## Likelihood ratio test=7.07 on 1 df, p=0.007848
## n= 76, number of events= 58
```

```
#Confidence intervals for each coefficient:
confint(cox0,level=.95)
```

```
##          2.5 %      97.5 %
## sex -1.419082 -0.2563841
```

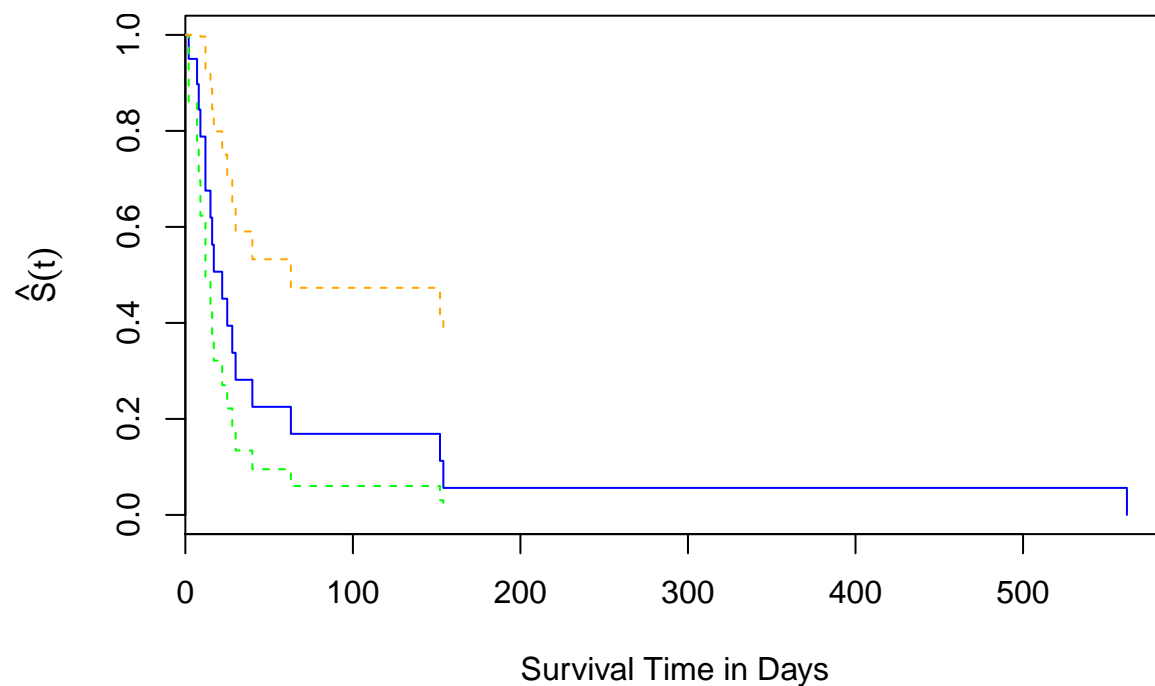
```
#Confidence intervals for the hazard ratios:
exp(confint(cox0,level=.95))
```

```
##          2.5 %      97.5 %
## sex 0.241936 0.7738447
```

The coefficient of kidney\$sex is -0.8377, and the hazard ratio is 0.4327. From above, we could know p-value is 0.00474 which is < 0.05. Therefore, we reject the H0. Hence, there is significant difference between the hazard proportion of two sexes. Basically, female hazard proportion (probabilities of going a month without an infection) is 0.8377 lower than male hazard proportion, and the 95% confidence interval for it would be (0.241936,0.7738447)

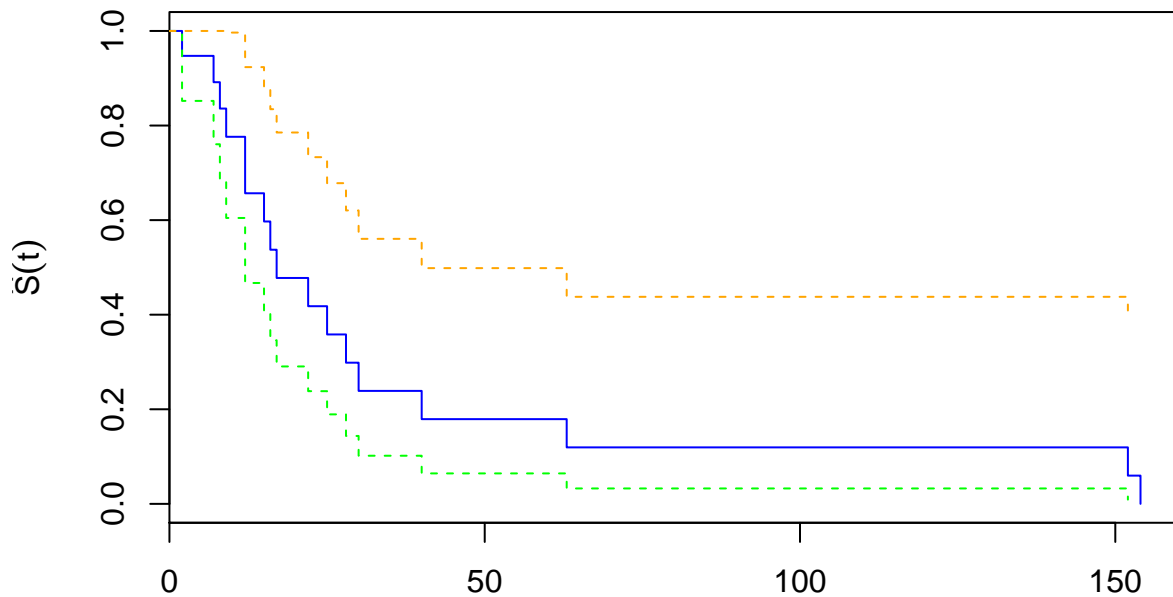
(e)

```
# first plot the KM estimate for male
par(mar = c(5,5,4,2))
plot(survfit(kid.surv ~ 1,
  subset = (kidney$sex == 1)),col = c("blue","green","orange"),
  xlab = "Survival Time in Days",
  ylab = expression(hat(S)(t)))
```



```
# Then Remove the ob in row 42
kidney1 <- kidney[-c(42),]
plot(survfit(Surv(kidney1$time,kidney1$status)~1,
```

```
subset = (kidney1$sex==1)), col = c("blue","green","orange"),
xlab = "Survival Time in Days",
ylab = expression(hat(S)(t)))
```



Survival Time in Days

From

the plots above, we could know the observation in row 42 is an outlier, and removing it would make the survival plot more appropriate.

2

```
data(mgus)
attach(mgus)
```

```
## The following objects are masked from kidney:
##
##   age, id, sex
```

```
mgus.time <- mgus$futime
mgus.cns <- mgus$death
mgus.surv <- Surv(mgus.time,mgus.cns)
```

(a)

Set up a Hypothesis Test: H_0 : There is a significant difference between the two sexes vs. H_a : there is no significant difference between the two sexes

```
#include 'sex'
cox0 <- coxph(mgus.surv~sex, data = mgus)
cox0
```

```
## Call:
## coxph(formula = mgus.surv ~ sex, data = mgus)
##
##           coef exp(coef) se(coef)      z      p
## sexmale 0.3385    1.4029   0.1360  2.489 0.0128
```

```
##
## Likelihood ratio test=6.28 on 1 df, p=0.01224
## n= 241, number of events= 225
```

As shown above, p-value is 0.0128, which is smaller than .05, meaning that we could conclude that there is a significant difference between the two sexes

(b)

H0: There is no significant difference between two sexes vs. Ha: There is a significant difference between two sexes Use Coxph function to find the p-value:

```
cox2 <- coxph(mgus.surv~age+alb+creat+hgb+mspike
              +sex, data=mgus)
cox2
```

```
## Call:
## coxph(formula = mgus.surv ~ age + alb + creat + hgb + mspike +
##       sex, data = mgus)
##
##               coef exp(coef) se(coef)      z      p
## age           0.070350  1.072884  0.008554  8.225 < 2e-16
## alb          -0.258449  0.772249  0.205974 -1.255 0.20957
## creat         0.405267  1.499702  0.147102  2.755 0.00587
## hgb          -0.106833  0.898676  0.060577 -1.764 0.07780
## mspike        0.010634  1.010691  0.199070  0.053 0.95740
## sexmale       0.205519  1.228162  0.165020  1.245 0.21298
##
## Likelihood ratio test=97.17 on 6 df, p=< 2.2e-16
## n= 176, number of events= 165
## (65 observations deleted due to missingness)
```

the p-value is 0.21298, which is greater than .05, so there is no statistically significant difference between 2 sexes.

(c)

In part (a), we did not exclude covariates to decrease their effect; however, we use the covariates age, alb, creat, hgb, and mspike to control for differences in the groups. Hence, we found that there is no significant difference between the sexes after using the covariates to control for differences.

(d)

Since from part b that p-value of age and creat are smaller than 0.05, we set age and creat as covariates:

```
cox4 <- coxph(mgus.surv~age+creat, data=mgus)
cox4
```

```
## Call:
## coxph(formula = mgus.surv ~ age + creat, data = mgus)
##
##               coef exp(coef) se(coef)      z      p
## age           0.075207  1.078107  0.008129  9.251 < 2e-16
## creat         0.489196  1.631005  0.132307  3.697 0.000218
##
## Likelihood ratio test=105.8 on 2 df, p=< 2.2e-16
## n= 198, number of events= 184
```

```
## (43 observations deleted due to missingness)
```

As presented above, p-value is $<2.2e-16 < 0.05$, which means there is significant difference between two sexes with the covariate age and creat

3

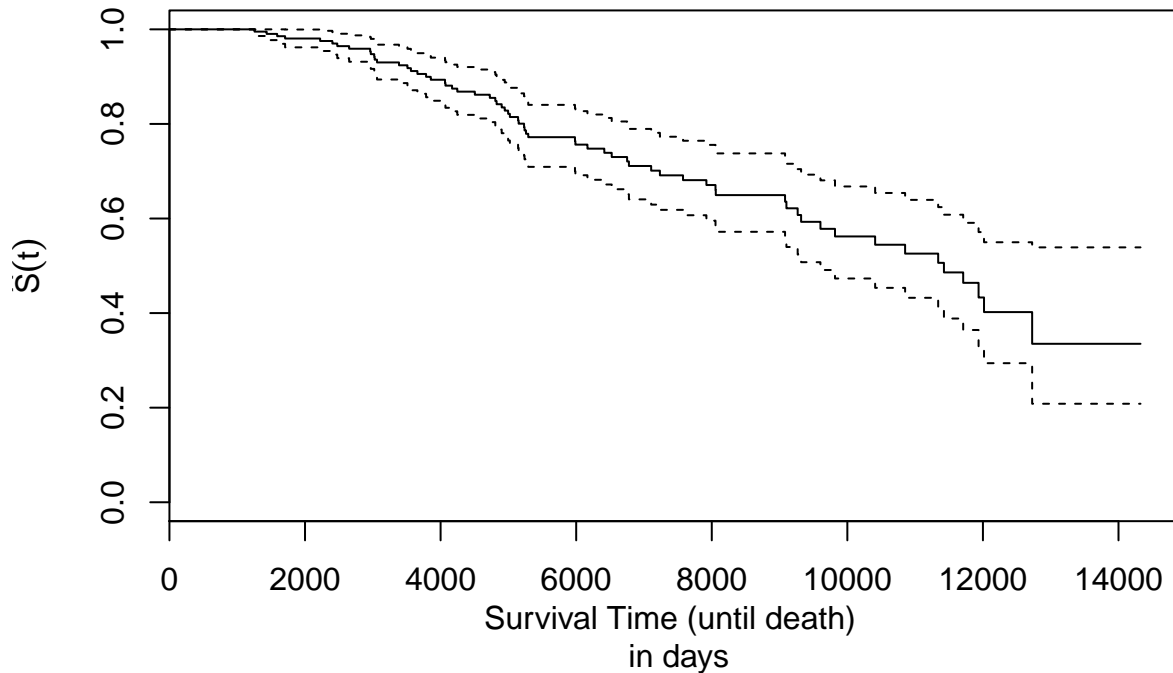
(a)

```
## The following is another solution which I have no idea if it is right or not
## Since the plot the following gives is different from the one in the next
## chunk. If the solution in the next chunk is not right, Please consider this ## solution. Also, the p
## changed.
```

```
## nmgus.a <- mgus
##nmgus.time <- for (i in seq(1:nrow(nmgus.a))){
##   if (is.na(nmgus.a[i,6])==TRUE){
##     nmgus.a[i,6]<-mgus$futime[i]
##   }
##}
##nmgus.b <- mgus
##nmgus.status <- for (i in seq(1:nrow(nmgus.b))) {
##   if (is.na(nmgus.b[i,6])==TRUE){
##     nmgus.b[i,6] <- 0
##   }else{
##     nmgus.b[i,6] <- 1
##   }
##}
##nmgus.a$pctime
##nmgus.b$pctime
##nmgus.surv = Surv(nmgus.a$pctime,nmgus.b$pctime)
##nmgus.fit <- survfit(nmgus.surv~1)
##plot(nmgus.fit, main = "Kaplan-Meier Curves \n for the time until onset of serious blood diseases",
##      xlab = "Survival Time (until death) \n in days",
##      ylab = expression(hat(S)(t)))
```

```
nmgus=mgus
for (i in seq(1:nrow(nmgus))){
  if (is.na(nmgus[i,6]) == TRUE){
    nmgus[i,6] <- 0
  }else{
    nmgus[i,6] <- 1
  }
}
nmgus.surv <- Surv(nmgus$futime,nmgus$pctime)
nmgus.fit <- survfit(nmgus.surv~1)
plot(nmgus.fit, main = "Kaplan-Meier Curves \n for the time until onset of serious blood diseases",
      xlab = "Survival Time (until death) \n in days",
      ylab = expression(hat(S)(t)))
```

Kaplan–Meier Curves for the time until onset of serious blood diseases



(b)

Use Coxph function to see if mspike has an effect on the time until a further disease is present:

```
nmgus <- mgus
cox6 <- coxph(nmgus.surv~nmgus$mspike,data=nmgus)
cox6
```

```
## Call:
## coxph(formula = nmgus.surv ~ nmgus$mspike, data = nmgus)
##
##               coef exp(coef) se(coef)      z      p
## nmgus$mspike -0.5947    0.5517  0.3224 -1.844 0.0651
##
## Likelihood ratio test=3.48 on 1 df, p=0.06206
## n= 241, number of events= 64
```

As shown above, we know that $p\text{-value} = 0.06206 > 0.05$. Therefore, we can conclude that mspike has no effect on the time until a further disease is present, and mspike is not significant.

(c)

```
#include 'mspike'
cox7 <- coxph(nmgus.surv~age+mspike+sex+alb+creat+hgb, data=nmgus)
#not include 'mspike'
cox8 <- coxph(nmgus.surv~age+sex+alb+creat+hgb, data=nmgus)

#Compute the Likelihood Ratio
lrt3 <- 2*(cox7$loglik[2]-cox8$loglik[2])
lrt3
```

```
## [1] 3.209515
```

```
#Approximate p with the Chi-squared distribution  
pchisq(lrt3,df=length(coef(cox7))-length(coef(cox8)),lower.tail=FALSE)
```

```
## [1] 0.07321119
```

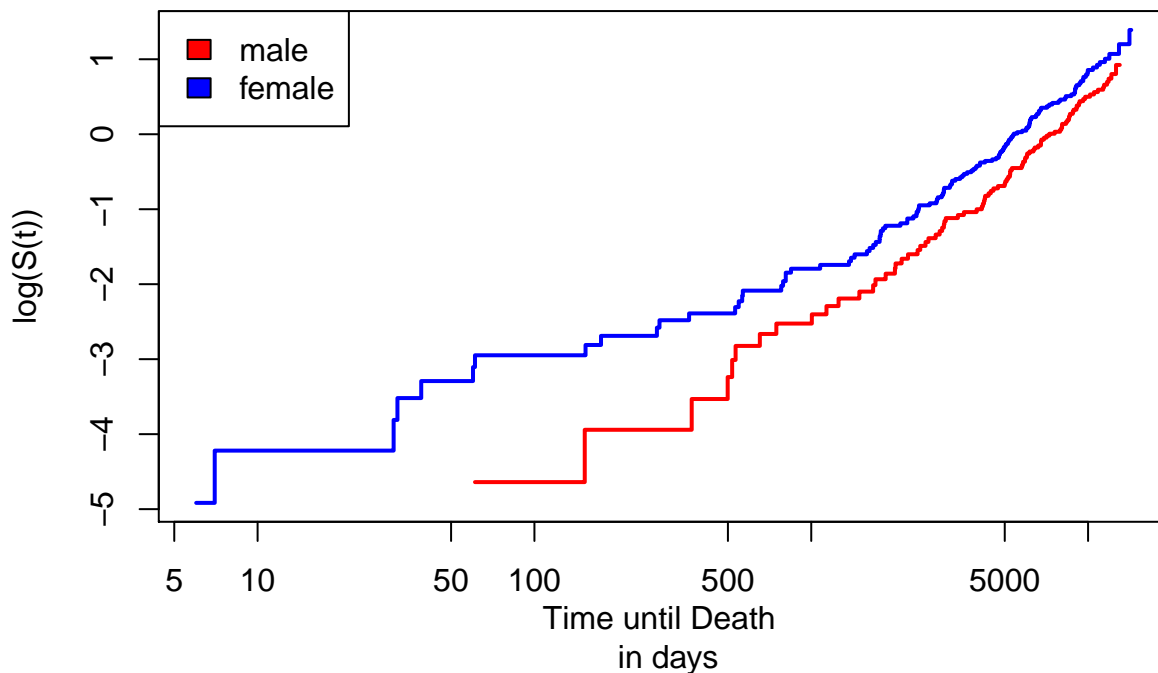
As shown above, p-value is 0.0732, which is > 0.05 . Therefore, we could conclude that mspike has no effect on the time until a further disease is present.

4

(a)

```
mgus.surv <- Surv(mgus.time,mgus.cns)  
plot(survfit(mgus.surv~mgus$sex),lwd=2,col=c(2,4),  
fun="cloglog",  
main = 'The -log-log Graphs of the Estimated Survival Functions \n for Male and Female Subjects',  
xlab="Time until Death \n in days",ylab="log(S(t))")  
legend('topleft',c("male","female"),fill = c("red","blue"))
```

**The -log-log Graphs of the Estimated Survival Functions
for Male and Female Subjects**



is no evidence that PH model is not appropriate because the curves do not cross each other.

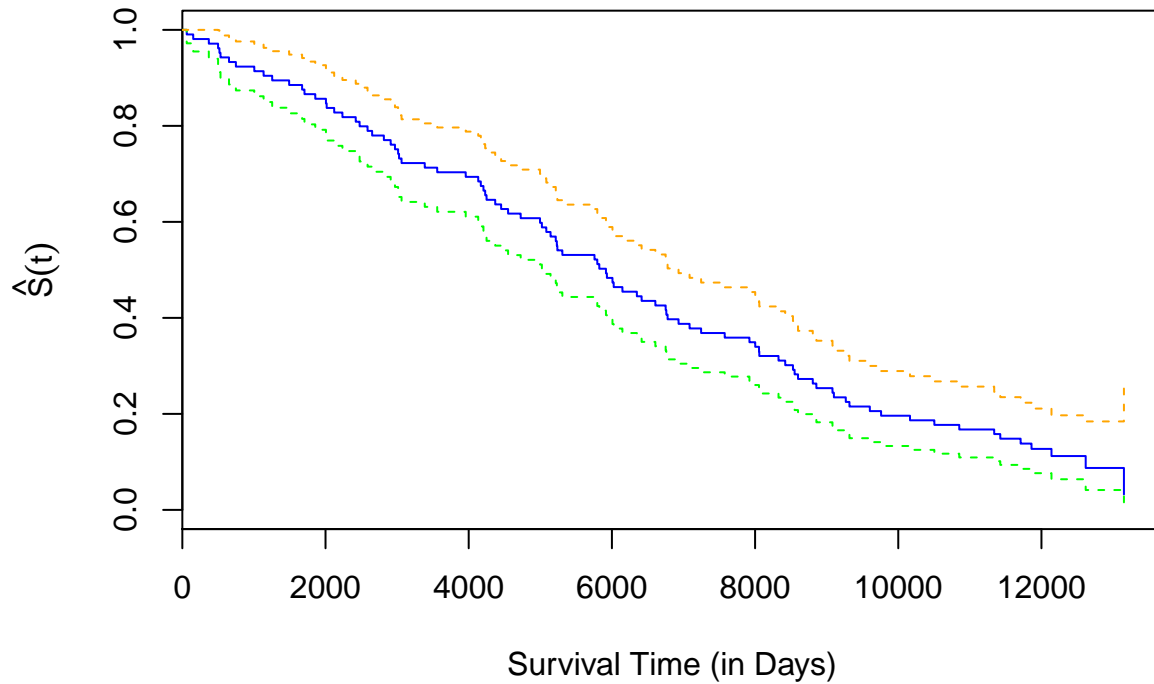
(b)

```
cox_1 <- coxph(mgus.surv ~ sex, subset = (mgus$sex == "female"), data = mgus)  
par(mar = c(5,5,4,2))  
plot(survfit(cox_1),  
main = "CoxPH Model for Female Subjects",  
xlab = "Survival Time (in Days)",
```



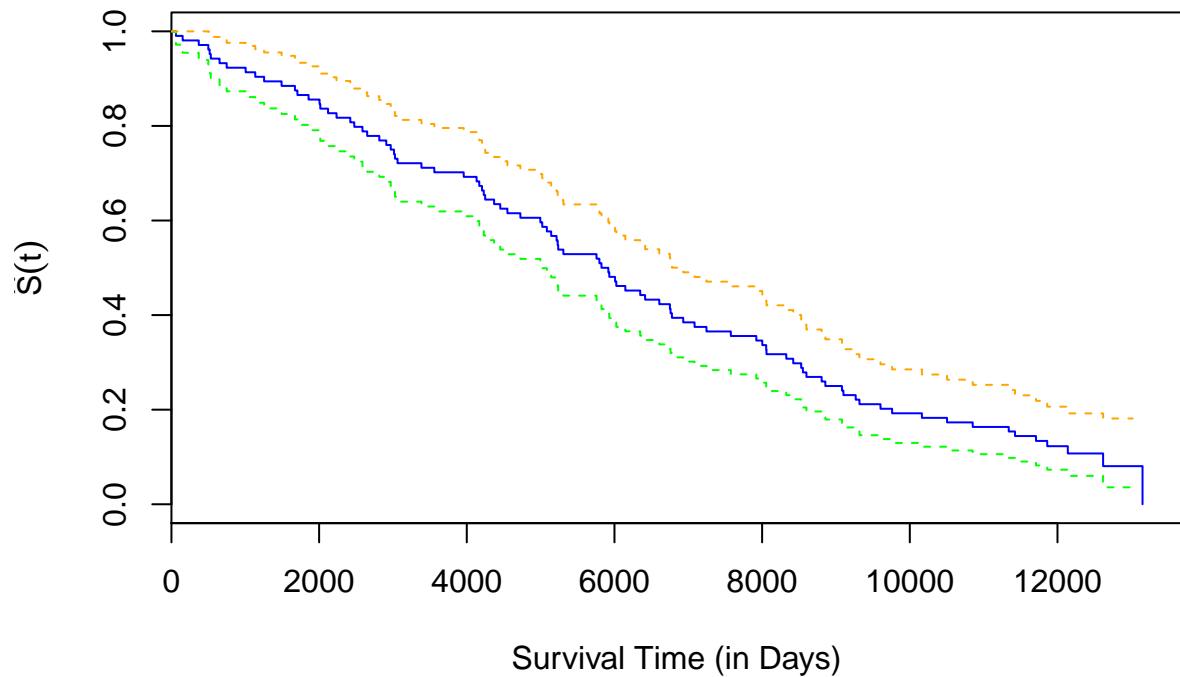
```
ylab = expression(hat(S)(t)),
col = c("blue", "green", "orange"))
```

CoxPH Model for Female Subjects



```
plot(survfit(mgus.surv~1, subset = (mgus$sex == "female")),
     main = "Kaplan-Meier Curves \n for Female Subjects",
     xlab = "Survival Time (in Days)",
     ylab = expression(hat(S)(t)),
     col = c("blue", "green", "orange"))
```

Kaplan–Meier Curves for Female Subjects



##

(c)

```
z = coxph(mgus.surv ~ sex, data = mgus)
cox.zph(z)
```

```
##          rho chisq    p
## sexmale -0.0833  1.53 0.216
```

From the result above, we could know p-value is 0.216 which is > 0.05 ; therefore, the model has no statistically significant difference with the proportional hazards model. Hence, we are justified in using the proportional hazards assumption in our modeling of the effect of sex, and since the result is not significant, we could conclude that the Cox PH Model is appropriate.