

PSTAT 175, LAB C

Kevin Ayala

November 08, 2018

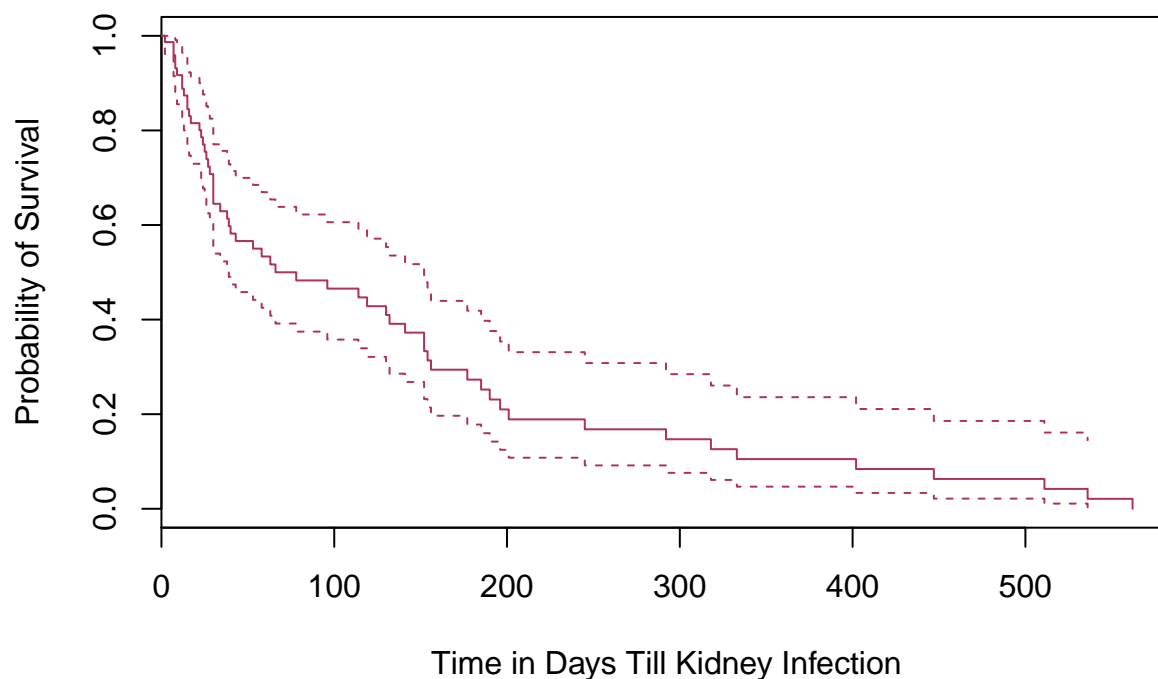
Question 1 Part A

?kidney

```
## No documentation for 'kidney' in specified packages and libraries:  
## you could try '??kidney'
```

```
library(survival)  
data("kidney")  
kidney.km <- survfit(Surv(kidney$time, kidney$status)~1)  
plot(kidney.km, xlab="Time in Days Till Kidney Infection", ylab="Probability of Survival",  
     main="Kidney Kaplan-Meier Plot", col= "maroon")
```

Kidney Kaplan-Meier Plot



Part B

```
log.rank.testkidney <- survdiff(Surv(kidney$time,kidney$status)~sex,data=kidney)  
log.rank.testkidney
```

```
## Call:
```

```
## survdiff(formula = Surv(kidney$time, kidney$status) ~ sex, data = kidney)
```

```
##
```

```
##      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
```

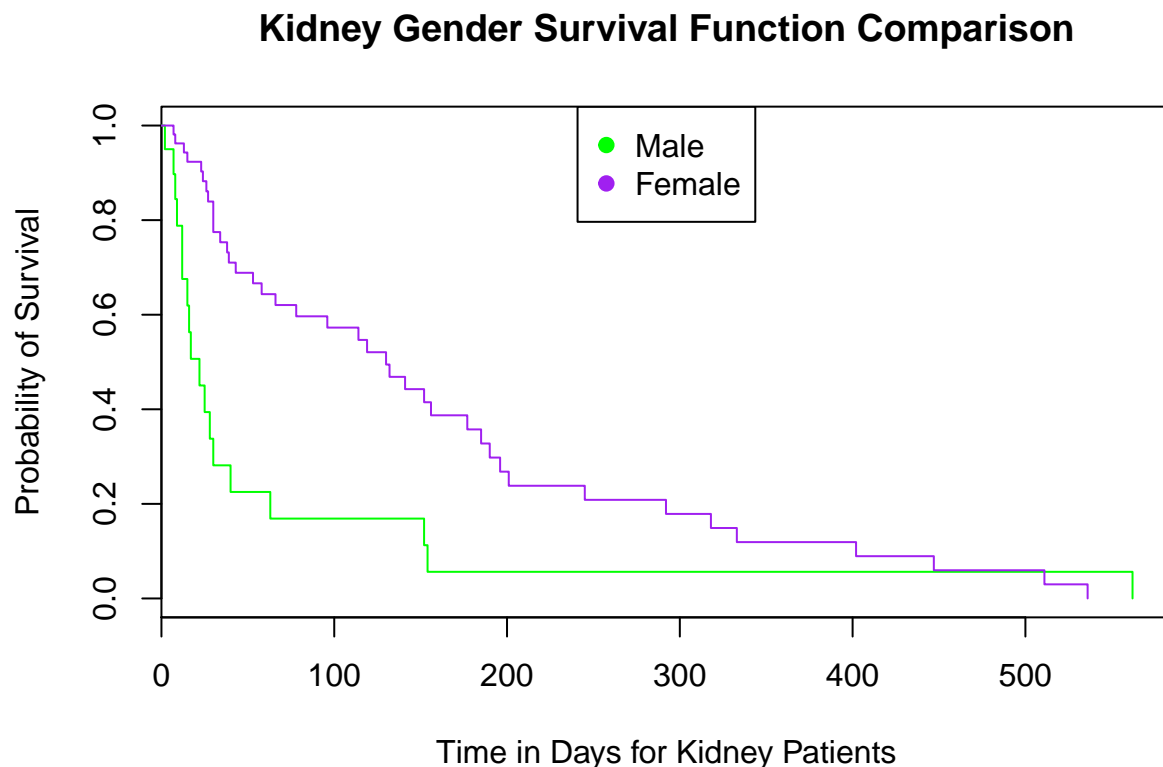
```
##
```

```
## Chisq= 8.3 on 1 degrees of freedom, p= 0.004
```

According to our log rank test, at an observed p value of .004, there is significant evidence to reject the null hypothesis of $H_0 : S_1(t) = S_2(t)$, and conclude the alternative $H_a : S_1(t) \neq S_2(t)$. This means that the survival rates till kidney infection is different for women and men.

Part C

```
kidney.km.sex = survfit(Surv(kidney$time,kidney$status)~sex,data=kidney)
plot(kidney.km.sex, xlab="Time in Days for Kidney Patients",
     ylab="Probability of Survival",
     main="Kidney Gender Survival Function Comparison",
     col=c("green","purple"))
legend("top",legend=c("Male","Female"), col=c("green","purple"), pch=rep(19,2))
```



Part D

```
kidney.coxph <- coxph(Surv(kidney$time,kidney$status)~sex,data=kidney)
kidney.coxph
```

```
## Call:
## coxph(formula = Surv(kidney$time, kidney$status) ~ sex, data = kidney)
##
##      coef exp(coef) se(coef)      z      p
## sex -0.838    0.433    0.297 -2.82 0.0047
##
## Likelihood ratio test=7.07 on 1 df, p=0.008
## n= 76, number of events= 58
exp(confint(kidney.coxph,level=.95))
```

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

We observe a 95% confidence interval for the estimate parameter which is between (.2419, .7738) for sex. Our hazard proportion by the cox proportion regression model is .433, this means that the survival rate is decreasing as time in days moves along for both sexes. In a month, the survival rate till a kidney infection has decreased and the patient will likely suffer a kidney infection compared till today.

Part E

```
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

kidney.male <- filter(kidney, sex == 1)
kidney.male.km<-survfit(Surv(kidney.male$time,kidney.male$status)~1)
summary(kidney.km.sex)

## Call: survfit(formula = Surv(kidney$time, kidney$status) ~ sex, data = kidney)
##
##               sex=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    2      20       1   0.9500  0.0487    0.85913    1.000
##    7      18       1   0.8972  0.0689    0.77183    1.000
##    8      17       1   0.8444  0.0826    0.69707    1.000
##    9      15       1   0.7881  0.0944    0.62328    0.997
##   12      14       2   0.6756  0.1094    0.49178    0.928
##   15      12       1   0.6193  0.1139    0.43186    0.888
##   16      11       1   0.5630  0.1166    0.37511    0.845
##   17      10       1   0.5067  0.1178    0.32128    0.799
##   22       9       1   0.4504  0.1174    0.27024    0.751
##   25       8       1   0.3941  0.1154    0.22197    0.700
##   28       7       1   0.3378  0.1118    0.17655    0.646
##   30       6       1   0.2815  0.1064    0.13417    0.591
##   40       5       1   0.2252  0.0989    0.09521    0.533
##   63       4       1   0.1689  0.0888    0.06029    0.473
##  152       3       1   0.1126  0.0749    0.03055    0.415
##  154       2       1   0.0563  0.0547    0.00839    0.378
##  562       1       1   0.0000    NaN          NA          NA
##
##               sex=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    7      53       1   0.9811  0.0187    0.94518    1.000
##    8      52       1   0.9623  0.0262    0.91231    1.000
##   13      50       1   0.9430  0.0320    0.88243    1.000
##   15      48       1   0.9234  0.0368    0.85393    0.998
##   23      45       1   0.9029  0.0413    0.82536    0.988
##   24      44       1   0.8823  0.0452    0.79804    0.976
##   26      41       1   0.8608  0.0490    0.77001    0.962
##   27      40       1   0.8393  0.0523    0.74289    0.948
##   30      39       3   0.7747  0.0601    0.66550    0.902
```

##	34	36	1	0.7532	0.0621	0.64076	0.885
##	38	35	1	0.7317	0.0640	0.61645	0.868
##	39	34	1	0.7102	0.0656	0.59253	0.851
##	43	33	1	0.6887	0.0671	0.56899	0.833
##	53	31	1	0.6664	0.0685	0.54486	0.815
##	58	29	1	0.6435	0.0699	0.52010	0.796
##	66	28	1	0.6205	0.0711	0.49573	0.777
##	78	26	1	0.5966	0.0722	0.47060	0.756
##	96	25	1	0.5727	0.0732	0.44589	0.736
##	114	22	1	0.5467	0.0743	0.41883	0.714
##	119	21	1	0.5207	0.0752	0.39230	0.691
##	130	20	1	0.4946	0.0758	0.36628	0.668
##	132	19	1	0.4686	0.0762	0.34076	0.644
##	141	18	1	0.4426	0.0763	0.31574	0.620
##	152	16	1	0.4149	0.0763	0.28930	0.595
##	156	15	1	0.3873	0.0761	0.26347	0.569
##	177	13	1	0.3575	0.0759	0.23584	0.542
##	185	12	1	0.3277	0.0752	0.20904	0.514
##	190	11	1	0.2979	0.0740	0.18308	0.485
##	196	10	1	0.2681	0.0723	0.15799	0.455
##	201	9	1	0.2383	0.0702	0.13382	0.424
##	245	8	1	0.2085	0.0674	0.11064	0.393
##	292	7	1	0.1787	0.0640	0.08856	0.361
##	318	6	1	0.1489	0.0599	0.06772	0.328
##	333	5	1	0.1192	0.0548	0.04836	0.294
##	402	4	1	0.0894	0.0485	0.03082	0.259
##	447	3	1	0.0596	0.0405	0.01573	0.226
##	511	2	1	0.0298	0.0292	0.00436	0.204
##	536	1	1	0.0000	NaN	NA	NA

Observation 42 (obs 42 is under sex =1 when time is 562) seems to have a survival hazard rate of 0% nor a confidence interval for his survival rate, this suggests he is an outlier. If we remove him for the kaplan meier then the model is free from a random subject effect.

Question 2 Part A

```
library(survival)
data("mgus")
log.rank.testmgus <- coxph(Surv(futime,death)~sex,data=mgus)
log.rank.testmgus
```

```
## Call:
## coxph(formula = Surv(futime, death) ~ sex, data = mgus)
##
##           coef exp(coef) se(coef)      z      p
## sexmale 0.339      1.403    0.136  2.49 0.013
##
## Likelihood ratio test=6.28 on 1 df, p=0.01
## n= 241, number of events= 225
```

We observe a pvalue of .01 from our log rank test, therefore we have significant evidence to reject the null hypotheses of $H_0 : S_1(t) = S_2(t)$ for the group of men and women, we thus conclude that the survival rates for both groups of sex to be different.

Part B

```
coxphtestmgus2 <- coxph(Surv(futime,death)~age+alb+creat+hgb+
                        mspike+sex,data=mgus)
coxphtestmgus3 <- coxph(Surv(futime,death)~age+alb+creat+hgb+
                        mspike,data=mgus)
loglikelihoodtest <- 2*(coxphtestmgus2$loglik[2]-coxphtestmgus3$loglik[2])
loglikelihoodtest
```

```
## [1] 1.562232
```

```
pchisq(loglikelihoodtest,df=1,lower.tail=FALSE)
```

```
## [1] 0.2113388
```

```
#loglikelihood test lets us use chisquare for p-value
```

We find evidence from the cox proportional hazard model above that the survival ratio between men and women is not different when we conduct a loglikelihood test with p-value of .2113, thus we fail to reject the null hypothesis of $H_0 : S_1(t) = S_2(t)$ between sexes.

Part C.

We explain that the two different answers in part a and b are because of the fact that the cox proportional hazard model can use multiple predictors and is a semi-parametric test. Because we use multiple predictors, we take into account continuous data as well as binary. For the Kaplan-Meier approach, we have to acknowledge that this approach is a non-parametric test and does not account for covariates.

Part D.

```
library(car)
```

```
## Loading required package: carData
```

```
##
```

```
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      recode
```

```
Anova(coxphtestmgus2)
```

```
## Analysis of Deviance Table (Type II tests)
```

```
##      LR Chisq Df Pr(>Chisq)
```

```
## age      73.195  1  < 2e-16 ***
```

```
## alb       1.566  1  0.21077
```

```
## creat     5.477  1  0.01927 *
```

```
## hgb       3.073  1  0.07962 .
```

```
## mspike     0.003  1  0.95741
```

```
## sex       1.562  1  0.21134
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We should use the covariates age and creat, based on an analysis of variance. We keep only these covariates because at a 95% level of significance, we find that only these two covariates play a significant role in the cox model. We reject all covariates with p-value's greater than .05 since they play no significant role.

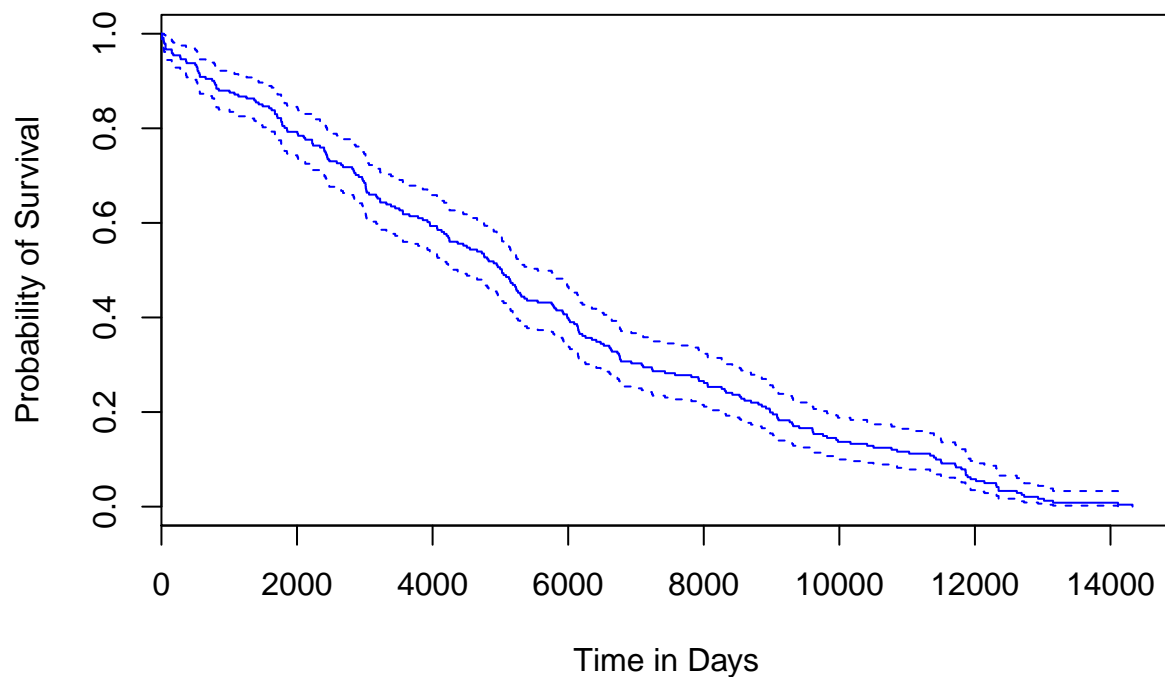
Question 3

Part A

```
#this code imputates NA values under pctime with corresponding futime.
mgus$pctime[is.na(mgus$pctime)]<- as.character(mgus$futime[is.na(mgus$pctime)])

newmgus<-survfit(Surv(as.numeric(mgus$futime),mgus$pctime,death))~1, data=mgus)
plot(newmgus, main="Kaplan-Meier Estimate with Imputed pctime",
      xlab="Time in Days", ylab="Probability of Survival" ,col="blue")
```

Kaplan-Meier Estimate with Imputed pctime



Part B

```
mgus.coxph <- coxph(Surv(mgus$futime,mgus$death)~mspike,data=mgus)
mgus.coxph

## Call:
## coxph(formula = Surv(mgus$futime, mgus$death) ~ mspike, data = mgus)
##
##           coef exp(coef) se(coef)  z    p
## mspike -0.167    0.846    0.168 -1 0.32
##
## Likelihood ratio test=1 on 1 df, p=0.3
## n= 241, number of events= 225
```

We observe a p-value of .3 for the cox model, we thus fail to reject the null hypothesis of $H_0 : S_1(t) = S_2(t)$ for mspike groups. Thus we conclude that mspike does not have an effect on time till a further disease is present.

Part C

```
mgus.coxph2 <- coxph(Surv(mgus$futime,mgus$death)~sex+age+alb+creat+
                    hgb+mspike,data=mgus)
mgus.coxph2
```

```
## Call:
```

```
## coxph(formula = Surv(mgus$futime, mgus$death) ~ sex + age + alb +
##      creat + hgb + mspike, data = mgus)
##
##           coef exp(coef) se(coef)      z      p
## sexmale  0.20552   1.22816  0.16502   1.25 0.2130
## age      0.07035   1.07288  0.00855   8.22 <2e-16
## alb     -0.25845   0.77225  0.20597  -1.25 0.2096
## creat    0.40527   1.49970  0.14710   2.76 0.0059
## hgb     -0.10683   0.89868  0.06058  -1.76 0.0778
## mspike   0.01063   1.01069  0.19907   0.05 0.9574
##
## Likelihood ratio test=97.17 on 6 df, p=<2e-16
## n= 176, number of events= 165
##      (65 observations deleted due to missingness)
mgus.coxph3 <- coxph(Surv(mgus$futime,mgus$death)~sex+age+alb+creat+
                    hgb, data=mgus)
mgus.coxph3
```

```
## Call:
## coxph(formula = Surv(mgus$futime, mgus$death) ~ sex + age + alb +
##      creat + hgb, data = mgus)
##
##           coef exp(coef) se(coef)      z      p
## sexmale  0.20472   1.22718  0.16431   1.25 0.2128
## age      0.07035   1.07288  0.00856   8.22 <2e-16
## alb     -0.25609   0.77407  0.20120  -1.27 0.2031
## creat    0.40571   1.50036  0.14672   2.77 0.0057
## hgb     -0.10708   0.89846  0.06041  -1.77 0.0763
##
## Likelihood ratio test=97.17 on 5 df, p=<2e-16
## n= 176, number of events= 165
##      (65 observations deleted due to missingness)
loglikelihood2 <- 2*(mgus.coxph2$loglik[2]-mgus.coxph3$loglik[2])
loglikelihood2
```

```
## [1] 0.00285201
pchisq(loglikelihood2,df=1,lower.tail=FALSE)
```

```
## [1] 0.9574099
```

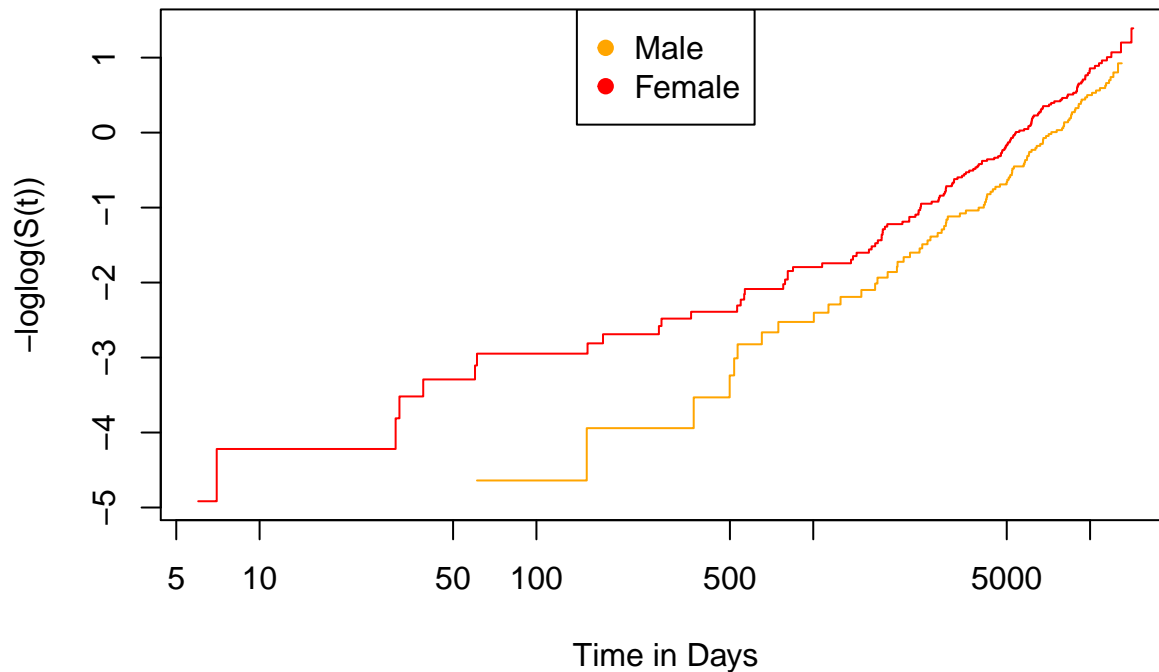
Considering the added covariates on mspike, we observe a p-value of .957 after doing a loglikelihood test which is approximately a chi square test with degrees of freedom =1. With our high p-value, we find no evidence to reject the null hypothesis and conclude $H_0 : S_1(t) = S_2(t)$. This further backs up our conclusion from part b.

Question 4

Part A

```
mgus.omega<- survfit(Surv(futime,death)~sex,data=mgus)
plot(mgus.omega, xlab="Time in Days", ylab="-loglog(S(t))",
main="-loglog(S(t)) Graph", col=c("orange","red"), fun="cloglog")
legend("top",legend=c("Male","Female"), col=c("orange","red"), pch=rep(19,2))
```

-loglog(S(t)) Graph



The model is appropriate since both graphs seem nice and parallel to each other, indicating the proportional hazard assumption is appropriate.

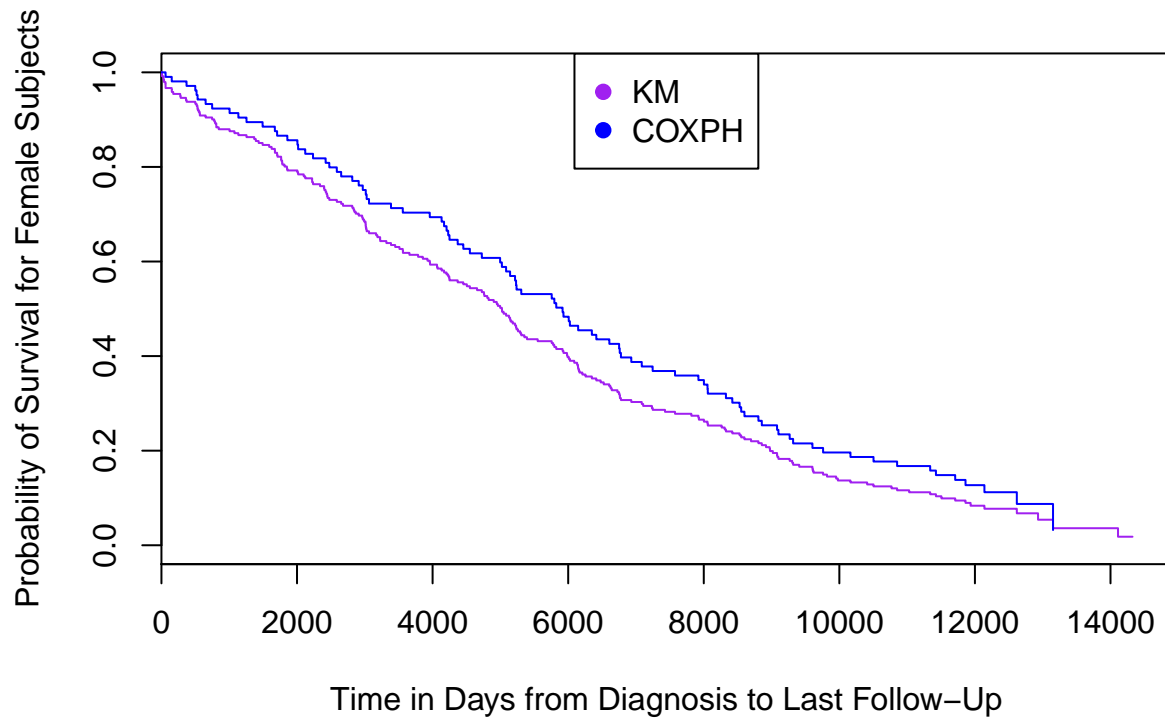
Part B

```
female.mgus <- filter(mgus, sex=="female")
female.mgus.coxph <- coxph(Surv(female.mgus$futime,female.mgus$death)~1)
female.mgus.coxph

## Call:  coxph(formula = Surv(female.mgus$futime, female.mgus$death) ~
##      1)
##
## Null model
##   log likelihood= -362.949
##   n= 104

female.mgus.km <- survfit(Surv(mgus$futime, mgus$death)~1,data=mgus)
plot(female.mgus.km, conf.int=FALSE, col="purple",
     xlab = "Time in Days from Diagnosis to Last Follow-Up",
     ylab = "Probability of Survival for Female Subjects",
     main= "KM and CoxPH Comparison for Female Subjects")
lines(survfit(female.mgus.coxph), conf.int=FALSE, col="blue")
legend("top",legend=c("KM","COXPH"), col=c("purple","blue"),pch=rep(19,2))
```


KM and CoxPH Comparison for Female Subjects



Yes the model seems to be appropriate since the KM estimate and Cox Proportional Hazard Model seem to be similar. If we were to plot confidence interval, then we would see crossing confidence intervals suggesting survival rates are possibly similar under Cox Model and Kaplan Meier estimate.

```
cox.zph(log.rank.testmgus, global = FALSE)
```

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

Yes we are justified in using the proportional hazard model of the effect of sex, we see under this test that the sex covariate has a p-value greater than .05. This means that the variable is not time dependent.