lab D

zhongyun zhang

2019/11/21

Problem 1

The data set that we used in lecture hern.txt contains information about treatment participation for heroin addicts in two clinics. In this problem, I want you to use likelihood ratio tests. ##a One of the covariates in the data, Prison, is a indicator of whether the subject had served prison time. Fit a Cox PH model to test whether prison has a significant effect on the time spent in the clinic.

```
hern <- read.table("~/Desktop/2019 Fall/PSTAT 175/lab4/Hern.txt", quote="\"", comment.char="")
library(survival)
head(hern)
```

```
##   ID Clinic Status Time Prison Dose
## 1  1      1      1  428      0   50
## 2  2      1      1  275      1   55
## 3  3      1      1  262      0   55
## 4  4      1      1  183      0   30
## 5  5      1      1  259      1   65
## 6  6      1      1  714      0   55
```

```
cox1 <- coxph(Surv(hern$Time,hern$Status)~Prison,data=hern)
summary(cox1)
```

```
## Call:
## coxph(formula = Surv(hern$Time, hern$Status) ~ Prison, data = hern)
##
##   n= 238, number of events= 150
##
##          coef exp(coef) se(coef)     z Pr(>|z|)
## Prison 0.1838    1.2018   0.1642 1.119    0.263
##
##        exp(coef) exp(-coef) lower .95 upper .95
## Prison     1.202     0.8321    0.8711     1.658
##
## Concordance= 0.536  (se = 0.023 )
## Likelihood ratio test= 1.25  on 1 df,   p=0.3
## Wald test            = 1.25  on 1 df,   p=0.3
## Score (logrank) test = 1.26  on 1 df,   p=0.3
```

The p value is p=0.3>0.05, which means covariate "Prison" has no significant effect on the survial time. Moreover, z for prison individually is 0.263>0.05, which also means "Prison" has no significant effect on the survival time.

b

The column Clinic tells us which clinic the subject was at. Repeat your test regarding the effect of prison time, but control for the possible confounding effect of the clinic.

```
cox2 <- coxph(Surv(hern$Time,hern$Status)~Prison+Clinic,data=hern)
cox3 <- coxph(Surv(hern$Time,hern$Status)~Clinic,data=hern)
summary(cox2)
```

```
## Call:
## coxph(formula = Surv(hern$Time, hern$Status) ~ Prison + Clinic,
##     data = hern)
##
##   n= 238, number of events= 150
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## Prison  0.2778    1.3203   0.1656  1.678   0.0933 .
## Clinic -1.1089    0.3299   0.2144 -5.173  2.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## Prison    1.3203     0.7574    0.9544    1.8263
## Clinic    0.3299     3.0309    0.2168    0.5022
##
## Concordance= 0.599  (se = 0.025 )
## Likelihood ratio test= 33.78  on 2 df,   p=5e-08
## Wald test            = 28.17  on 2 df,   p=8e-07
## Score (logrank) test = 30.51  on 2 df,   p=2e-07
```

```
lrt = 2*(cox2$loglik[2]-cox3$loglik[2])
pchisq(lrt,df=1,lower.tail = FALSE)
```

```
## [1] 0.09493244
```

The p value is p=0.0949>0.05, which clinic does not have significant effect on the survival time when controling the effect of clinic.

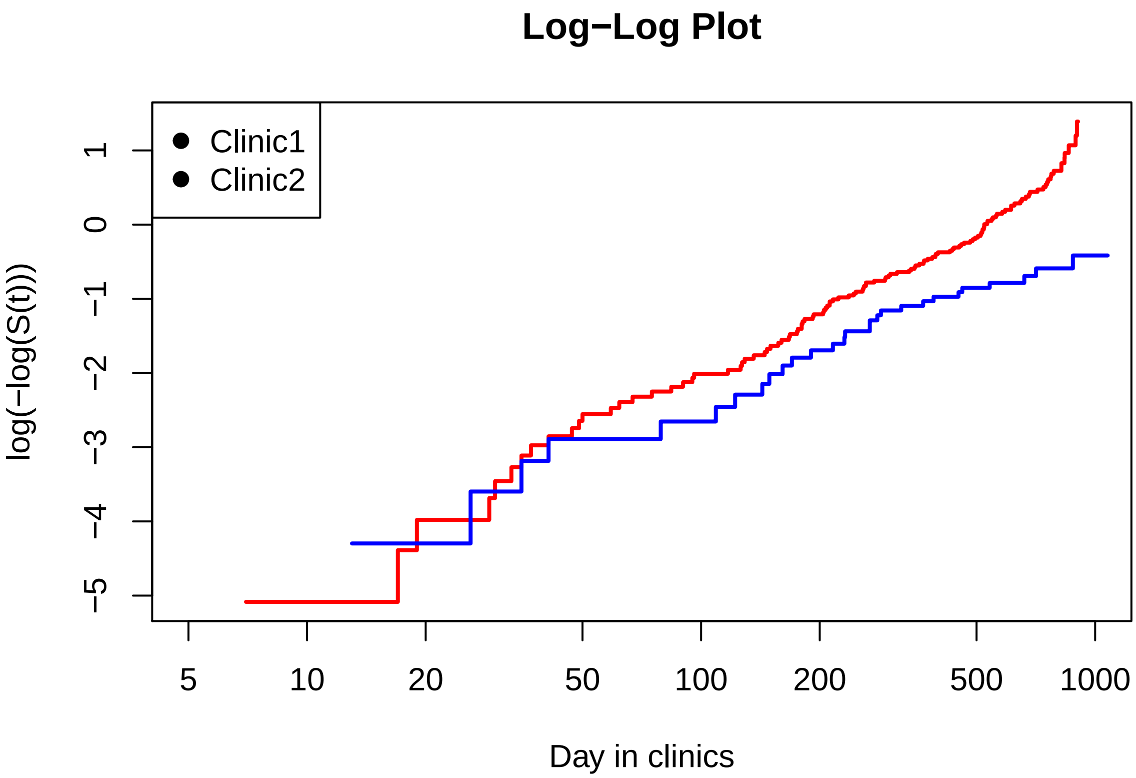
c

Use a complementary log-log plot to visualize whether the Cox PH assumption is appropriate for modeling the effect of the clinics. Describe you conclusions from looking at the plot.

```
par(mar=c(5,5,4,2))
plot(survfit(Surv(Time,Status)~Clinic,data=hern),col=c(2,4),lwd=2, xlim=c(5,1000),fun="cloglog",main="L
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): 1 x value <= 0 omitted
## from logarithmic plot
```

```
legend("topleft",legend=c("Clinic1","Clinic2"),pch=rep(19,2))
```



two curves cross over each other for patients who stay in the clinics less than 50. so the coxph assumption is not appropriate for modeling the effect of the clinics.

d

Model the effect of prison when you treat the clinics as stratified confounding variables. What do you conclude? Describe in your own words the difference between this model and the model in part (b).

```
cox4 = coxph(Surv(Time,Status)~Prison+strata(Clinic),data=hern)
anova(cox4)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(Time, Status)
## Terms added sequentially (first to last)
##
##         loglik  Chisq Df Pr(>|Chi|)
## NULL   -614.47
## Prison -612.48 3.9788  1    0.04608 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the likelihood ratio test, p-value is 0.04608, which means that Prison is significant while clinic is stratified. In c, because the model is without clinic stratified, it cannot fit the model, we should modify the covariate.

e

Perform a test to see if there is a significant interaction between the prison variable and the clinic variable. We still want to use a stratified model.What do you conclude? Explain what the interaction term means.

```
cox5 = coxph(Surv(Time,Status)~Prison*strata(Clinic),data=hern)
summary(cox5)
```

```
## Call:
## coxph(formula = Surv(Time, Status) ~ Prison * strata(Clinic),
##     data = hern)
##
##   n= 238, number of events= 150
##
##                                  coef exp(coef) se(coef)      z Pr(>|z|)
## Prison                         0.4055    1.5000   0.1862  2.177   0.0294 *
## Prison:strata(Clinic)Clinic=2 -0.3585    0.6987   0.4236 -0.846   0.3973
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                               exp(coef) exp(-coef) lower .95 upper .95
## Prison                           1.5000     0.6667    1.0413     2.161
## Prison:strata(Clinic)Clinic=2    0.6987     1.4312    0.3046     1.603
##
## Concordance= 0.544  (se = 0.023 )
## Likelihood ratio test= 4.7  on 2 df,   p=0.1
## Wald test            = 4.76  on 2 df,   p=0.09
## Score (logrank) test = 4.82  on 2 df,   p=0.09
```

```
anova(cox5)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(Time, Status)
## Terms added sequentially (first to last)
##
##                        loglik  Chisq Df Pr(>|Chi|)
## NULL                  -614.47
## Prison                -612.48 3.9788  1    0.04608 *
## Prison:strata(Clinic) -612.12 0.7183  1    0.39670
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is p=0.3967>0.05, which mean that there are no significant interaction between the prison variable and the clinic variable.

Problem 2

The data set retire has information on the life expectancy of individuals living in a senior care facility. We begin by modeling time column which is the survival time in months spent at the facility. The indicator column death will be used as our status variable. We would like to model the difference between men and women so there is a column gender which is 1 for men and 2 for women. ##a Use a Cox Proportional Hazards model to test whether there is a significant difference between men and women. Report the likelihood ratio statistic and the appropriate P value.

```
library(survival)
retire <- read.table("~/Desktop/2019 Fall/PSTAT 175/lab4/retire.txt", header=TRUE)
attach(retire)
head(retire)
```

```
##   obs death ageentry age time gender
## 1 272     0      733 870  137      2
## 2  67     0      746 804   58      2
## 3  50     1      748 804   56      2
## 4 451     1      751 777   26      1
```

```
## 5 455     1      759 781   22      1
## 6 192     0      760 897  137      2
```

```
cox6 <- coxph(Surv(retire$time,retire$death)~gender,data=retire)
summary(cox6)
```

```
## Call:
## coxph(formula = Surv(retire$time, retire$death) ~ gender, data = retire)
##
##   n= 462, number of events= 176
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## gender -0.4280    0.6518   0.1718 -2.492   0.0127 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## gender    0.6518      1.534    0.4655    0.9127
##
## Concordance= 0.54  (se = 0.017 )
## Likelihood ratio test= 5.78  on 1 df,   p=0.02
## Wald test            = 6.21  on 1 df,   p=0.01
## Score (logrank) test = 6.3  on 1 df,   p=0.01
```

The test statistic is 5.78, p value is p=0.02<0.05,we reject null hypothesis and conclude that gender has significant effect between men and women survival time.

b

Fit another model that adjusts for the confounding variable ageentry which gives the age in months of the subject when they entered the facility. Use the anova function to calculate the appropriate likelihood ratio test. Do you come to the same conclusion as in part (a)? How do you explain any difference?

```
cox7 <- coxph(Surv(retire$time,retire$death)~ageentry+gender,data=retire)
anova(cox7)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(retire$time, retire$death)
## Terms added sequentially (first to last)
##
##           loglik   Chisq Df Pr(>|Chi|)
## NULL     -974.89
## ageentry -952.39 44.9854  1  1.985e-11 ***
## gender   -950.15  4.4917  1    0.03406 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p value is p=0.03406<0.05, we reject the null hypothesis and conclude that there is significant difference between men and women when we control the effect of ageentry. WE get the same conclusion form part A an the difference is whether we control the effect of ageentry.

c

Fit a model with an interaction between age and gender. What do you conclude?

```
cox8 = coxph(Surv(time,death)~ageentry*gender,data=retire)
anova(cox8)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(time, death)
## Terms added sequentially (first to last)
##
##                    loglik   Chisq Df Pr(>|Chi|)
## NULL              -974.89
## ageentry          -952.39 44.9854  1  1.985e-11 ***
## gender            -950.15  4.4917  1    0.03406 *
## ageentry:gender -949.58  1.1424  1    0.28515
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p value is p=0.028515<0.05, which intersection between ageetry and gender have significant effect on the survival time.

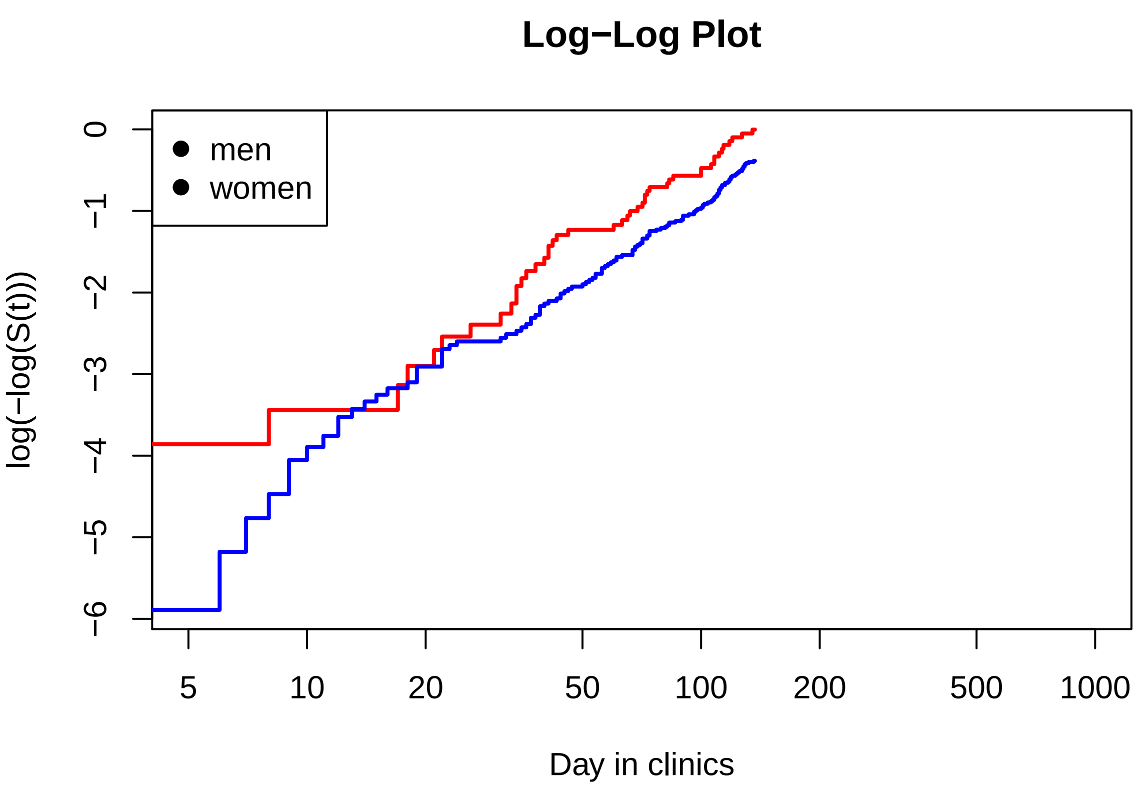
d

Plot complementary log-log plot comparing the effect of gender on the survival time. Do you think the proportional hazards assumption is reasonable for this model?

```r
par(mar=c(5,5,4,2))
plot(survfit(Surv(time,death)~retire$gender,data=retire),col=c(2,4),lwd=2, xlim=c(5,1000),fun="cloglog"
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): 1 x value <= 0 omitted
## from logarithmic plot
```

```r
legend("topleft",legend=c("men","women"),pch=rep(19,2))
```



two curves cross over so the coxph assumption is not appropriate for modeling the effect of the gender.

e

Explain clearly why we chose to use ageentry as our covariate and not age which is the age of the subject when the event occurred. ageentry is the age people enter the clinic, and we analysis the time people stay and their death time. Age is the time they dead, we cannot analysis anything from this. We choose ageentry because it is the starting time of the event and will not change when time change. It is more likely a stable variable. The age is changing during the experiment and is not stable.

Problem 3

Using the same retire data set, I want to fit a Generalized Cox Model where the effect of gender is different before and after 48 months. ##a Use the survSplit function to construct a new data frame with additional rows that split the time variable into before and after 48 months.

```
retire2 <- survSplit(Surv(time, death) ~., retire, cut=48,zero = -1, episode ="timegroup")
head(retire2)
```

```
##   obs ageentry age gender tstart time death timegroup
## 1 272      733 870      2     -1   48     0         1
## 2 272      733 870      2     48  137     0         2
## 3  67      746 804      2     -1   48     0         1
## 4  67      746 804      2     48   58     0         2
## 5  50      748 804      2     -1   48     0         1
## 6  50      748 804      2     48   56     1         2
```

b

Use coxph to model the effect of gender including a change of parameter at 48 months. Please include the age at entry if that is still appropriate. Use our likelihood ratio test to determine if gender is still significant in this model.

```
cox9 <- coxph(Surv(tstart, time, death)~ ageentry+ gender:timegroup, data = retire2)
anova(cox9)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(tstart, time, death)
## Terms added sequentially (first to last)
##
##                   loglik   Chisq Df Pr(>|Chi|)
## NULL             -974.89
## ageentry         -952.39 44.9854  1  1.985e-11 ***
## gender:timegroup -950.88  3.0314  1    0.08167 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this case z is 0.08167>0.05, which means gender is not significant in this model.

c

Give 95% confidence intervals for the hazard ratio for men before and after the 48 month cut off.

```
cox3c<-coxph(Surv(tstart, time, death)~ageentry+gender+gender*strata(timegroup),data=retire2)
cox3c
```

```
## Call:
## coxph(formula = Surv(tstart, time, death) ~ ageentry + gender +
##     gender * strata(timegroup), data = retire2)
##
```

```
##                                        coef exp(coef)  se(coef)      z
## ageentry                            0.007100  1.007126  0.001049  6.770
## gender                             -0.574733  0.562855  0.266615 -2.156
## gender:strata(timegroup)timegroup=2  0.332986  1.395127  0.350236  0.951
##                                          p
## ageentry                            1.28e-11
## gender                                0.0311
## gender:strata(timegroup)timegroup=2   0.3417
##
## Likelihood ratio test=50.38  on 3 df, p=6.638e-11
## n= 767, number of events= 176
```

Since the P value is 0.08167 and we fail to reject null hypothesis. we conclude that gender is not siginifcant in this model.

d

Would you conclude that it is important to consider a change in the effect of gender before and after 4 years in the retirement facility?

```
cox3d<-coxph(Surv(tstart, time, death)~gender:strata(timegroup),data=retire2)
summary(cox3d)
```

```
## Call:
## coxph(formula = Surv(tstart, time, death) ~ gender:strata(timegroup),
##     data = retire2)
##
##   n= 767, number of events= 176
##
##                                        coef exp(coef) se(coef)      z
## gender:strata(timegroup)timegroup=1 -0.6516    0.5212   0.2664 -2.446
## gender:strata(timegroup)timegroup=2 -0.2783    0.7571   0.2271 -1.226
##                                     Pr(>|z|)
## gender:strata(timegroup)timegroup=1   0.0144 *
## gender:strata(timegroup)timegroup=2   0.2204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                     exp(coef) exp(-coef) lower .95
## gender:strata(timegroup)timegroup=1    0.5212      1.919    0.3092
## gender:strata(timegroup)timegroup=2    0.7571      1.321    0.4851
##                                     upper .95
## gender:strata(timegroup)timegroup=1    0.8785
## gender:strata(timegroup)timegroup=2    1.1815
##
## Concordance= 0.54  (se = 0.017 )
## Likelihood ratio test= 6.91  on 2 df,   p=0.03
## Wald test            = 7.49  on 2 df,   p=0.02
## Score (logrank) test = 7.71  on 2 df,   p=0.02
```

In time group 1 which is before 4 years. There is significant difference between gender, but in the later 4 years, there is no siginifcant changed between gender. It is significant to consider before and after 4 years.