Lab D

Mujie Wang

11/22/2019

```r
library(survival)
library(MASS)
hern1 <- read.table("/Users/wangmujie/Desktop/hern.txt",header = TRUE)
retire <- read.table("/Users/wangmujie/Desktop/retire.txt", header = TRUE)
```

1(a)

```r
hern1.surv <- Surv(hern1$Time, hern1$Status)
hern1.pr <- coxph(hern1.surv ~ hern1$Prison)
summary(hern1.pr)
```

```
## Call:
## coxph(formula = hern1.surv ~ hern1$Prison)
##
##   n= 238, number of events= 150
##
##                coef exp(coef) se(coef)     z Pr(>|z|)
## hern1$Prison 0.1838    1.2018   0.1642 1.119    0.263
##
##              exp(coef) exp(-coef) lower .95 upper .95
## hern1$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 of prison is 0.263 which is greater than alpha $= 0.05$, there is no significant effect.

1(b)

```r
hern1.pr.cl <- coxph(hern1.surv ~ Clinic+Prison, data = hern1)
hern1.cl <- coxph(hern1.surv ~ Clinic, data = hern1)
lrt.prison <- 2*(hern1.pr.cl$loglik[2]-hern1.cl$loglik[2])
pchisq(lrt.prison, df= 1, lower.tail = FALSE)
```

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

Since the p-value is 0.09493 which is gereater than a $= 0.05$, which means that the prison time is not a significant given the clinic is included in the model.

1(c)

```r
hern1.fit.cl <- survfit(hern1.surv ~ Clinic, data = hern1)

plot(hern1.fit.cl,fun="cloglog", main="Log-log plot of Heroin Addicts from Two Clinics",
     xlab = "days in treatment program", ylab = "log(-log S(t))",
     lwd=2, col=2:1, mark.time=TRUE, mark = 20)
```

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

```
legend("topleft",inset = 0.10,legend=c("Clinic 1", "Clinic2"),
       pch=15,col=2:1)
```

Log−log plot of Heroin Addicts from Two Clinics
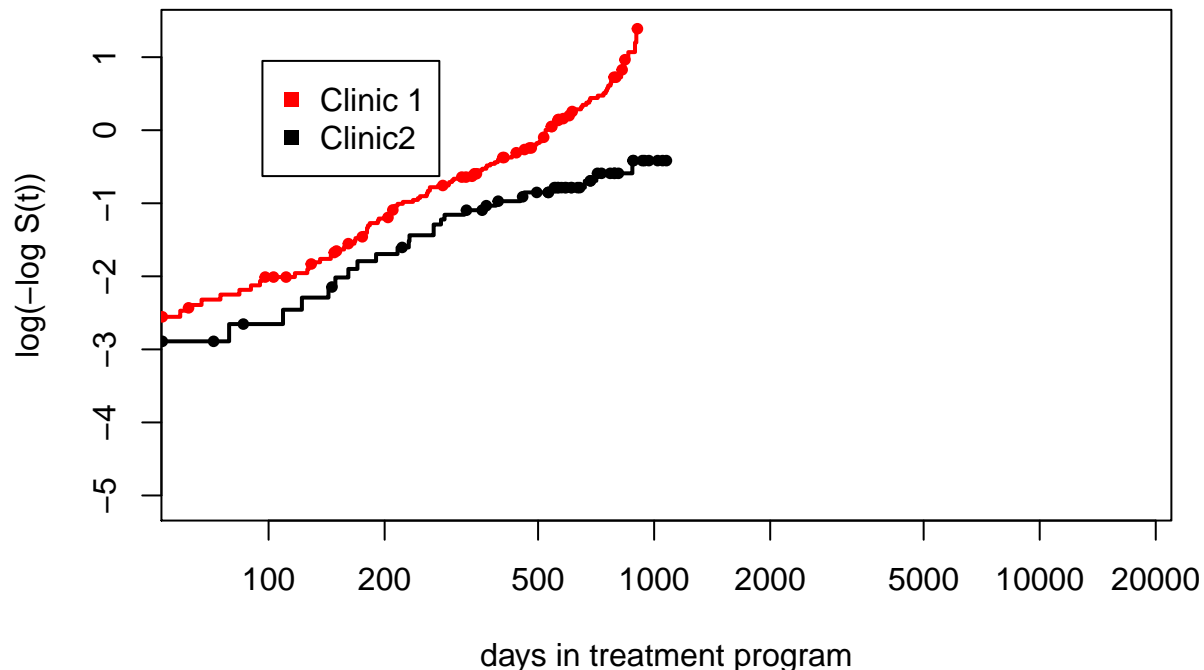
In the plot there two line represent the clinics, and are two crossing point. At the same time, the lines are diverged as the time of the treatment program increases.

1(d)

```
hern1.pr.stcl <- coxph(hern1.surv ~ Prison+ strata(Clinic), data = hern1)
summary(hern1.pr.stcl)
```

```
## Call:
## coxph(formula = hern1.surv ~ Prison + strata(Clinic), data = hern1)
##
##   n= 238, number of events= 150
##
##           coef exp(coef) se(coef)     z Pr(>|z|)
## Prison 0.3359    1.3992   0.1675 2.005    0.045 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## Prison     1.399     0.7147     1.008     1.943
##
## Concordance= 0.544  (se = 0.023 )
## Likelihood ratio test= 3.98  on 1 df,   p=0.05
## Wald test            = 4.02  on 1 df,   p=0.04
## Score (logrank) test = 4.05  on 1 df,   p=0.04
```

The p-value is 0.045, which is less than a = 0.05. It mean that the prison time have effect. In part 1b, we assime that the baseline hazard rate is the same for all clinics, but this model we use strata(Clinic), so it

```

make the constant hazard ratio more significant with Prison.

1(e)

```
hern1.pr_stCl <- coxph(hern1.surv ~ Prison*strata(Clinic), data = hern1)
summary(hern1.pr_stCl)
```

```
## Call:
## coxph(formula = hern1.surv ~ Prison * strata(Clinic), data = hern1)
##
##   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
```

```
lrt.prC <- 2*(hern1.pr_stCl$loglik[2]-hern1.pr.stcl$loglik[2])
pchisq(lrt.prC, df=1,lower.tail = FALSE)
```

```
## [1] 0.3966961
```

```
anova(hern1.pr_stCl,hern1.pr.stcl)
```

```
## Analysis of Deviance Table
##  Cox model: response is  hern1.surv
##  Model 1: ~ Prison * strata(Clinic)
##  Model 2: ~ Prison + strata(Clinic)
##    loglik  Chisq Df P(>|Chi|)
## 1 -612.12
## 2 -612.48 0.7183  1    0.3967
```

This is a test about significant interaction between the prison variable and the clinic variable, and the p-value is 0.3966961 which is greater than 0.05, means that they are not significant. The interaction term means that the subject the prison time or not and which clinic they go to, and see how long they stays in the treatment program for the heroin addiction. It can also help use to decide if one variable depended on the other.

2(a)

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

```
## Call:
## coxph(formula = retire.surv ~ 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 likelihood ratio statistic is 5.78 and the p-value is 0.016 which is less then 0.05, so that the survival time between 2 genders has a significant different.

2(b)

```
retire.a <- coxph(retire.surv ~ ageentry, data = retire)
retire.g.a <- coxph(retire.surv ~ gender + ageentry, data = retire)
anova(retire.a,retire.g.a )
```

```
## Analysis of Deviance Table
##  Cox model: response is  retire.surv
##  Model 1: ~ ageentry
##  Model 2: ~ gender + ageentry
##    loglik  Chisq Df P(>|Chi|)
## 1 -952.39
## 2 -950.15 4.4917  1   0.03406 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is 0.03406 which is less then 0.05, so there is a significant difference in the survival time between different gender (men and women). However, compare to part a, they have a different model fitting. In part a, it did not control the ageentry; In part b, it did the age will keep constant after they entered the facility.

2(c)

```
retire.g_a <- coxph(retire.surv ~ gender + age + gender*age, data = retire)
lrt.retire <- 2*(retire.g_a$loglik[2] - retire.ge$loglik[2])
pchisq(lrt.retire, df=1, lower.tail = FALSE)
```

```
## [1] 0.001362656
```

```
retire.g_ag <- coxph(retire.surv ~ gender + ageentry + gender*ageentry, data = retire)
lrt.retire_ag <- 2*(retire.g_ag$loglik[2] - retire.ge$loglik[2])
pchisq(lrt.retire_ag, df=1, lower.tail = FALSE)
```

```
## [1] 2.137116e-11
```

For age: The interaction between age and gender is signifocant or not is showed why using the likelihood ratio test and the interaction is significant because the p-value is 0.0014 which is less than 0.05. For agentry: The interaction between age and gender is signifocant or not is showed why using the likelihood ratio test and the interaction is significant because the p-value is 2.137116e-11 which is less than 0.05.

2(d)

```
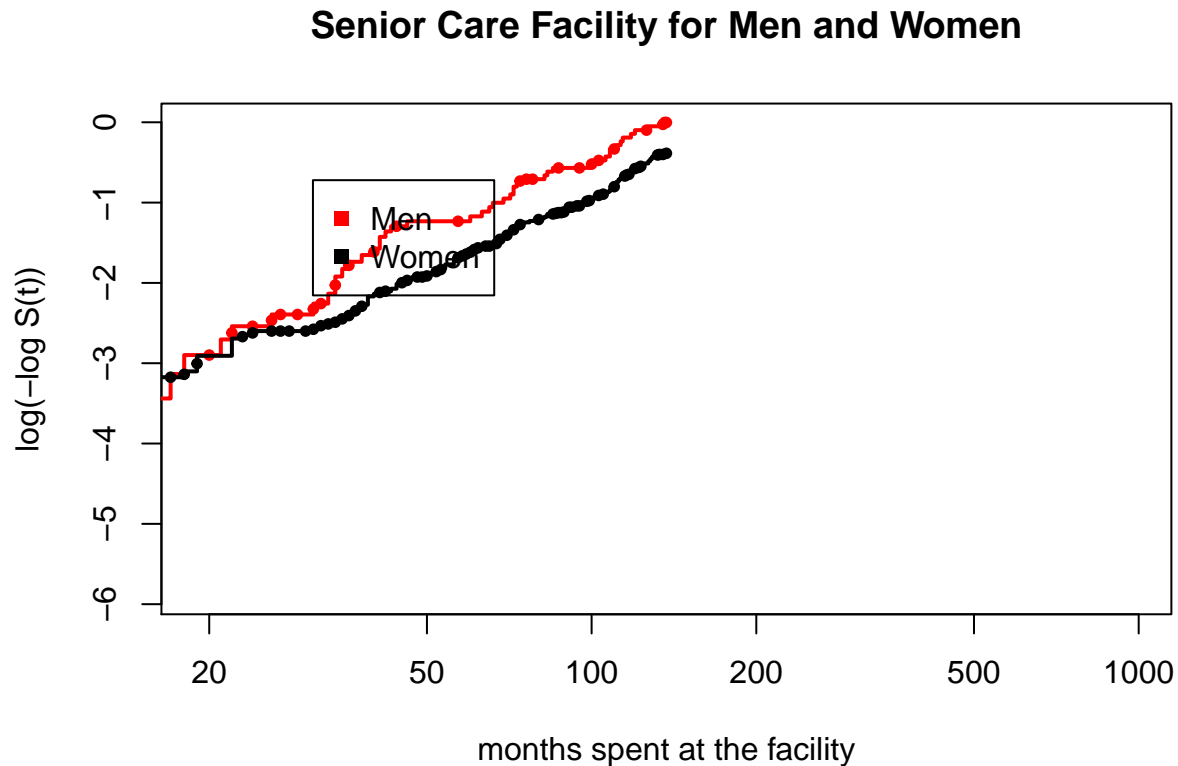retire.fit <- survfit(retire.surv ~ gender, data = retire)
plot(retire.fit,fun="cloglog", main="Senior Care Facility for Men and Women",
```

```
    xlab = "months spent at the facility", ylab = "log(-log S(t))", lwd=2,
    col=2:1, mark.time=TRUE, mark = 20)
```

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

```
legend("topleft",inset = 0.15,legend=c("Men", "Women"),pch=15,col=2:1)
```

Senior Care Facility for Men and Women

months spent at the facility

The crossing lines show that the proportional hazards assumption is not reasonable for the model. This plot does not satisfy the constant distance between the differentce in the log(-log(s(t))) that we want so have.

2(e) We use ageentry as our covariate and not age which is the age of the subject when the event occurred because we don???t want to assume that younger individuals have a longer survival time than the older. It is because there factors are affecting the survival time such as illness they have before. We also want to focus on how the time of individual stay in the senior care affect their survivial time.

3(a)

```
retire_3 <- retire
retire_3$time[retire_3$time == 0] = 0.05
retire_3.split <- survSplit(Surv(time,death)~., episode = "episode",
                            cut = 48, id = "sub.id", data = retire_3)
retire_3.split[1:5, ]
```

```
##   obs ageentry age gender sub.id tstart time death episode
## 1 272      733 870      2      1      0   48     0       1
## 2 272      733 870      2      1     48  137     0       2
## 3  67      746 804      2      2      0   48     0       1
## 4  67      746 804      2      2     48   58     0       2
## 5  50      748 804      2      3      0   48     0       1
```

3(b)

```
retire_3.splitfit <- coxph(Surv(tstart, time, death)~ gender:strata(episode),
                            data = retire_3.split)
retire_3.splitfit2 <- coxph(Surv(tstart, time, death)~ ageentry + gender:strata(episode),
                            data = retire_3.split)
lrt.retire_3split <- 2*(retire_3.splitfit2$loglik[2]-retire_3.splitfit$loglik[2])
pchisq(lrt.retire_3split, df=1, lower.tail = FALSE)
```

```
## [1] 4.31289e-11
```

```
summary(retire_3.splitfit2)
```

```
## Call:
## coxph(formula = Surv(tstart, time, death) ~ ageentry + gender:strata(episode),
##     data = retire_3.split)
##
##   n= 767, number of events= 176
##
##                                    coef exp(coef)  se(coef)      z
## ageentry                       0.007100  1.007126  0.001049  6.770
## gender:strata(episode)episode=1 -0.574733  0.562855  0.266615 -2.156
## gender:strata(episode)episode=2 -0.241748  0.785254  0.227129 -1.064
##                                Pr(>|z|)
## ageentry                       1.28e-11 ***
## gender:strata(episode)episode=1   0.0311 *
## gender:strata(episode)episode=2   0.2872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                exp(coef) exp(-coef) lower .95 upper .95
## ageentry                          1.0071     0.9929    1.0051    1.0092
## gender:strata(episode)episode=1   0.5629     1.7767    0.3338    0.9492
## gender:strata(episode)episode=2   0.7853     1.2735    0.5031    1.2256
##
## Concordance= 0.648  (se = 0.023 )
## Likelihood ratio test= 50.38  on 3 df,    p=7e-11
## Wald test            = 52.64  on 3 df,    p=2e-11
## Score (logrank) test = 53.84  on 3 df,    p=1e-11
```

```
anova(retire_3.splitfit,retire_3.splitfit2)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(tstart, time, death)
##  Model 1: ~ gender:strata(episode)
##  Model 2: ~ ageentry + gender:strata(episode)
##     loglik  Chisq Df P(>|Chi|)
## 1 -971.43
## 2 -949.70 43.467  1 4.313e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3(c)

```
ep1 <- subset(retire_3.split, retire_3.split$episode==1)
ep2 <- subset(retire_3.split, retire_3.split$episode==2)


ep1.coxph <- coxph(Surv(time,death)~gender, data = ep1)
```

```
summary(ep1.coxph)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ gender, data = ep1)
##
##   n= 462, number of events= 64
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## gender -0.6516    0.5212   0.2664 -2.446   0.0144 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## gender    0.5212      1.919    0.3092    0.8785
##
## Concordance= 0.559  (se = 0.029 )
## Likelihood ratio test= 5.49  on 1 df,   p=0.02
## Wald test            = 5.98  on 1 df,   p=0.01
## Score (logrank) test = 6.2  on 1 df,   p=0.01
```

```
ep2.coxph <- coxph(Surv(time,death)~gender, data = ep2)
summary(ep2.coxph)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ gender, data = ep2)
##
##   n= 305, number of events= 112
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## gender -0.2783    0.7571   0.2271 -1.226     0.22
##
##        exp(coef) exp(-coef) lower .95 upper .95
## gender    0.7571      1.321    0.4851     1.181
##
## Concordance= 0.521  (se = 0.019 )
## Likelihood ratio test= 1.43  on 1 df,   p=0.2
## Wald test            = 1.5  on 1 df,   p=0.2
## Score (logrank) test = 1.51  on 1 df,   p=0.2
```

The 95% CI of the hazard ratio for men before 48 months cutoff is [0.3092, 0.8785] and after 48 months cutoff is [0.4851, 1.181].

3(d) I conclude that it is important to consider a change in the effect of gender before and after 4 years in the retirement facility, in the 95% CI for the hazard ratio before the 4 years cutoff from part 3c the result is [0.3092, 0.8785], so we can conclude what the effect of gender is siginificat to the model. While the after 4 years cut off, the 95% CI is [0.4851, 1.181], which show that the effect of gender is not significant because 1 is inluded. I think that it maybe because people in different gender will have similar treatment so there are less different betern the hazard rate and survival time for different gender.