175 LAB D

Kevin Ayala

11/29/2018

Question 1 Part A.

```
library(survival)
library(KMsurv)

getwd()
```

```
## [1] "/Users/kevinlorenzoayala/Downloads"
```

```
hern.data<-read.table("/Users/kevinlorenzoayala/Downloads/hern.txt")
head(hern.data)
```

```
##   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
```

```
fittedmodel1 = coxph(Surv(Time,Status)~Prison,data=hern.data)
summary(fittedmodel1)
```

```
## Call:
## coxph(formula = Surv(Time, Status) ~ Prison, data = hern.data)
##
##   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 )
## Rsquare= 0.005   (max possible= 0.997 )
## 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 test reveals that we should fail to reject null with a p-value of .3. There exists no difference between both prisons under the null hypothesis. We can conclude both prisons share the same hazard rate.

Part B.

```
fittedmodel2 <- coxph(Surv(Time, Status)~Prison+Clinic, data = hern.data)
fittedmodel3 <- coxph(Surv(Time, Status)~Clinic, data = hern.data)


logratiotest<- 2*(fittedmodel2$loglik[2]-fittedmodel3$loglik[2])
pchisq(logratiotest, df=1, lower.tail=FALSE) #df is 1 due to groups
```

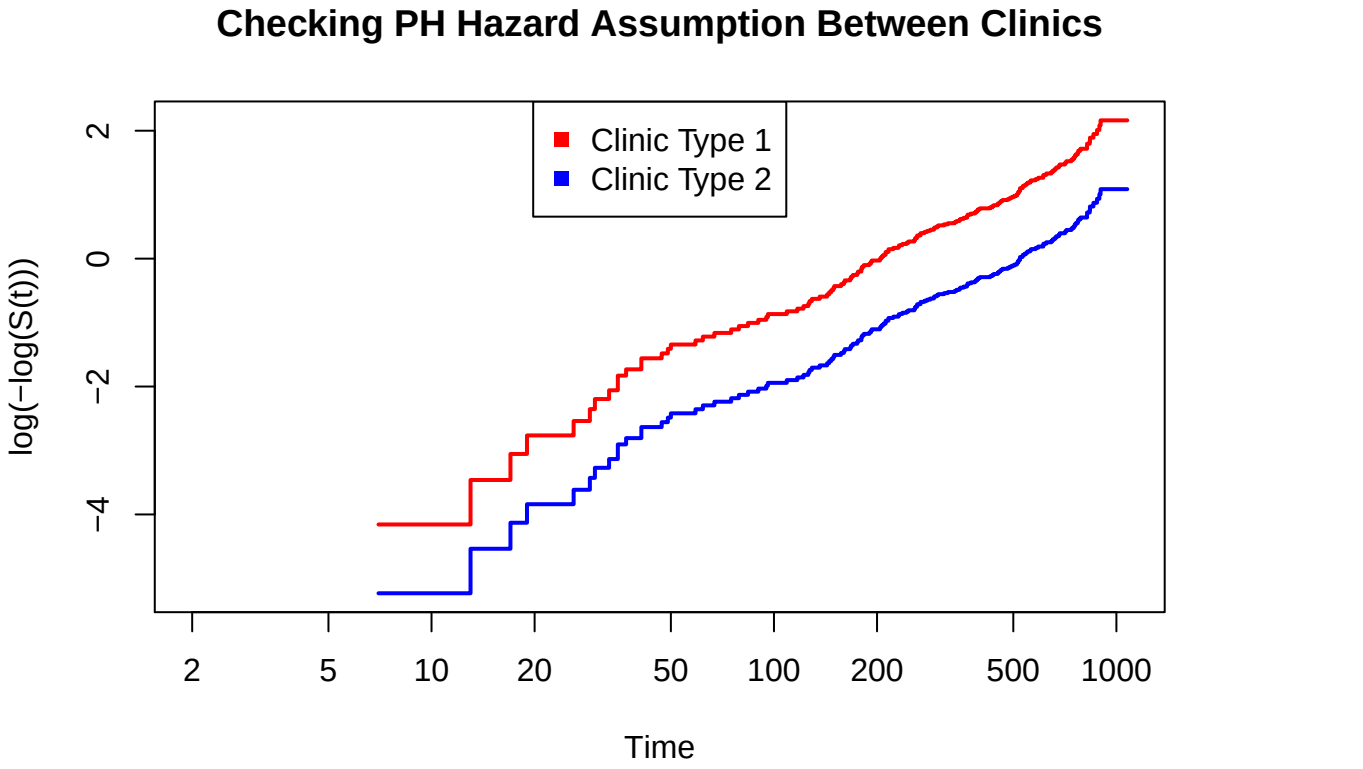
```
## [1] 0.09493244
```

We use model in question 1 to control for clinic for the log liklihood test. Under the log likelihood test, we observe a pvalue of .09. We fail to reject null based on this. We conclude that the effect of prison is not affected by the confounding variable clinic.

Part C.

```
fittedmodel3 <- coxph(Surv(Time,Status)~Clinic,data=hern.data)

plot(survfit(fittedmodel3, newdata=data.frame(Clinic=factor(c("1","2")))),
fun="cloglog",col= c("red","blue"),xlab="Time",ylab="log(-log(S(t)))",lwd=2, main =
  "Checking PH Hazard Assumption Between Clinics")
legend("top", legend = c("Clinic Type 1", "Clinic Type 2"), col = c("red","blue"), pch=rep(15,4))
```



We observe both loglog plots of both clinics are parallel to each and do not seem like they will cross, therefore on a visual check, we conclude the Cox PH assumption is accepted for the effect of clinics.

Part D.

```
stratisfiedmodel1 = coxph(Surv(Time,Status)~Prison+strata(Clinic), data=hern.data)
summary(stratisfiedmodel1)
```

```
## Call:
## coxph(formula = Surv(Time, Status) ~ Prison + strata(Clinic),
##     data = hern.data)
##
##   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 = 0.029 )
## Rsquare= 0.017   (max possible= 0.994 )
## 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
```

We should conlcude that the clinics share the same hazard ratios since the pvalue is .05, since the pvalue is not greater than our alpha level of .05 we reject the null hypothesis. The clinic variable when stratisfied is significant, meaning the stratisfication takes into account different groups under the clinic.

Part E.

```
#Part E

fittedmodelinter <- coxph(Surv(Time, Status)~Prison*strata(Clinic), data = hern.data)

lrt = 2*(fittedmodelinter$loglik[2]-stratisfiedmodel1$loglik[2])
pchisq(lrt,df=1,lower.tail = FALSE)
```

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

```
#for anova test we fit new cox model with prison and clinic interaction
anova(stratisfiedmodel1, fittedmodelinter)
```

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

```
#anova test confirms lrt outcome
```

We fail to reject the null with an observed p-value of .3967. We conclude the interaction term is not significant. Thus we dont add the interaction. We used an anova test to confirm our likelyhood result..

Problem 2 Part A.

```
#a
getwd()
```

```
## [1] "/Users/kevinlorenzoayala/Downloads"
```

```
retire.data<-read.table("/Users/kevinlorenzoayala/Downloads/Retire.txt", header = TRUE)
head(retire.data)
```

```
##   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
```

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

```
## Call:
## coxph(formula = Surv(time, death) ~ gender, data = retire.data)
##
##   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 = 0.016 )
## Rsquare= 0.012   (max possible= 0.985 )
## 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
```

We observe a likelyhood ration test of 5.78 with a corresponding p-value of .02, this means that we reject the null hypothesis that there is no difference between both gender and thus conclude that there is a difference in the hazard functions for both groups based on the data.

Part B.

```
coxphageentry<- coxph(Surv(time,death)~gender+ageentry, data=retire.data)
anova(coxphgender, coxphageentry)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(time, death)
##  Model 1: ~ gender
##  Model 2: ~ gender + ageentry
##    loglik  Chisq Df P(>|Chi|)
## 1 -972.00
## 2 -950.15 43.699  1 3.831e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#
logratiotest2<- 2*(coxphageentry$loglik[2]-coxphgender$loglik[2])
logratiotest2
```

```
## [1] 43.6985
```

```
pchisq(logratiotest2, df=1, lower.tail=FALSE)
```

```
## [1] 3.83065e-11
```

our log likilyhood ratio test is given by the anova when we take the difference of log like which is the difference of the log likelyhood between each model. Thus to get the log likelyhood we do 2*(difference of log likelyhood for each model) which results in 2*(-950.15 + 972) which gives us 43.7. We confirm this with the code under logratiotest2.

pvalue of 9.318827e-10, we reject null hypothesis and conclude age entry and gender plays an interaction. We conclude on top of having gender in data, we should keep ageentry in the final model.

Part C.

```
coxgenderretire<- coxph(Surv(time,death) ~ age*gender, data=retire.data)
summary(coxgenderretire)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age * gender, data = retire.data)
##
##   n= 462, number of events= 176
##
##                 coef exp(coef)  se(coef)      z Pr(>|z|)
## age        -0.011789  0.988280  0.004783 -2.465   0.0137 *
## gender     -5.426193  0.004400  2.619183 -2.072   0.0383 *
## age:gender  0.005012  1.005024  0.002644  1.895   0.0580 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##            exp(coef) exp(-coef) lower .95 upper .95
## age           0.9883      1.012 9.791e-01    0.9976
## gender        0.0044    227.282 2.594e-05    0.7463
## age:gender    1.0050      0.995 9.998e-01    1.0102
##
## Concordance= 0.575  (se = 0.023 )
## Rsquare= 0.034   (max possible= 0.985 )
## Likelihood ratio test= 16.03  on 3 df,   p=0.001
## Wald test            = 19.08  on 3 df,   p=3e-04
## Score (logrank) test = 19.62  on 3 df,   p=2e-04
```

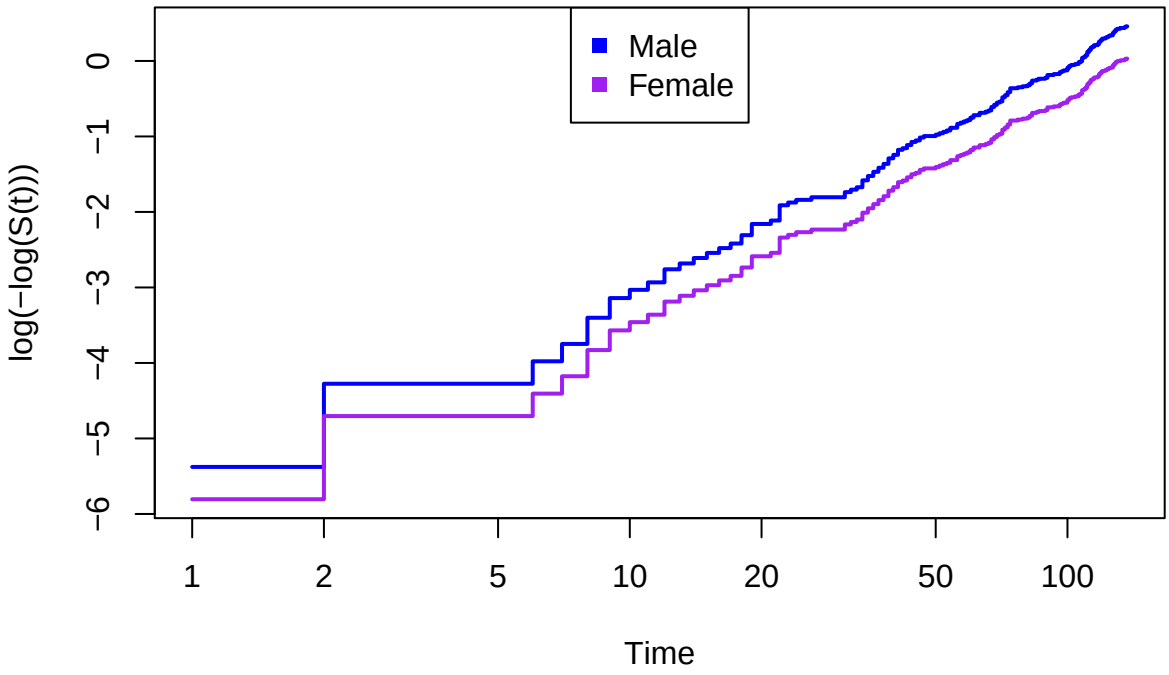
```
anova(coxphgender , coxgenderretire)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(time, death)
##  Model 1: ~ gender
##  Model 2: ~ age * gender
##    loglik  Chisq Df P(>|Chi|)
## 1 -972.00
## 2 -966.87 10.256  2  0.005929 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We observe a pvalue of 0.001, therefore we find evidence that age and gender play an interaction roll. We confirm this result with the anova code. We infer that although age and gender play an interaction, we find that due to occam's razor that ageentry might be better since it has a lower p-value and is more significant than age.

Part D.

```
gender.survival <- coxph(Surv(time,death)~gender, data= retire.data)
plot(survfit(gender.survival, newdata=data.frame(gender=factor(c("1","2")))),
fun="cloglog",col= c("blue","purple"),xlab="Time",ylab="log(-log(S(t)))",lwd=2, main =
  "Checking PH Hazard Assumption Between Genders")
legend("top", legend = c("Male", "Female"), col = c("blue","purple"), pch=rep(15,4))
```



Yes, we find our proportional hazard assumption as met, we find it appropriate that the model assumes this assumption. We come to this conclusion because the log log graph indicates no interaction between gender.

Part E.

We want to use ageentry because the p-value is more significant than age alone and since we don't want to have a overcomplicated model with two differnt age variables, by law of parsimony we should use ageentry as it is just more simpler.

Question 3 Part A .

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#a
head(retire.data)
```

```
##   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
```

```
retire.data$time[retire.data$time == 0] <- .1
#this code adds time value for time 0 observations so that the code for survsplit works
min(retire.data$time)
```

```
## [1] 0.1
```

```
newfit = survSplit(Surv(time, death) ~., data=retire.data , cut = c(48),episode ="timegroup", id ="sub.
newfit[1:10,c("sub.id","tstart","time","death","timegroup","age","gender")]
```

```
##    sub.id tstart time death timegroup age gender
## 1       1      0   48     0         1 870      2
## 2       1     48  137     0         2 870      2
## 3       2      0   48     0         1 804      2
## 4       2     48   58     0         2 804      2
## 5       3      0   48     0         1 804      2
## 6       3     48   56     1         2 804      2
## 7       4      0   26     1         1 777      1
## 8       5      0   22     1         1 781      1
## 9       6      0   48     0         1 897      2
## 10      6     48  137     0         2 897      2
```

Part B.

```
newmodelfit2 = coxph(Surv(tstart,time,death)~gender+ageentry, data=newfit)
newmodelfit2
```

```
## Call:
## coxph(formula = Surv(tstart, time, death) ~ gender + ageentry,
##     data = newfit)
##
##              coef exp(coef) se(coef)     z       p
## gender   -0.37591   0.68666  0.17191 -2.19   0.029
## ageentry  0.00712   1.00715  0.00105  6.79 1.1e-11
##
## Likelihood ratio test=49.48  on 2 df, p=2e-11
## n= 767, number of events= 176
```

```
newmodelfit3 = coxph(Surv(tstart,time,death)~gender:strata(timegroup)+ageentry,data=newfit)


anova(newmodelfit2,newmodelfit3)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(tstart, time, death)
##  Model 1: ~ gender + ageentry
##  Model 2: ~ gender:strata(timegroup) + ageentry
##    loglik  Chisq Df P(>|Chi|)
## 1 -950.15
## 2 -949.70 0.9005  1    0.3426
```

```
logratiotest3<- 2*(newmodelfit3$loglik[2]-newmodelfit2$loglik[2])
logratiotest3
```

```
## [1] 0.9005031
```

```
pchisq(logratiotest3, df=1, lower.tail=FALSE)
```

```
## [1] 0.3426469
```

We fail to reject the null hypothesis and conclude that gender is not significant in this model.

Part C.

```
#c
newfitmen <-filter(retire.data, time < 48 & gender ==1)
head(newfitmen)
```

```
##   obs death ageentry age time gender
## 1 451     1      751 777   26      1
## 2 455     1      759 781   22      1
## 3 391     0      817 843   26      1
## 4 449     1      836 876   40      1
## 5 431     0      846 866   20      1
## 6 436     1      847 893   46      1
```

```
coxfitmen <- coxph(Surv(time, death)~ageentry, data = newfitmen)
summary(coxfitmen)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ ageentry, data = newfitmen)
##
##   n= 41, number of events= 21
##
##              coef exp(coef) se(coef)     z Pr(>|z|)
## ageentry 0.001359  1.001360 0.003426 0.397    0.692
##
##          exp(coef) exp(-coef) lower .95 upper .95
## ageentry     1.001     0.9986    0.9947     1.008
##
## Concordance= 0.547  (se = 0.08 )
## Rsquare= 0.004   (max possible= 0.929 )
## Likelihood ratio test= 0.16  on 1 df,   p=0.7
## Wald test            = 0.16  on 1 df,   p=0.7
## Score (logrank) test = 0.16  on 1 df,   p=0.7
```

```
exp(confint(coxfitmen,level=.95))
```

```
##              2.5 %   97.5 %
## ageentry 0.9946578 1.008107
```

```
newfitmen2 <-filter(retire.data, time > 48 & gender ==1)
head(newfitmen2)
```

```
##   obs death ageentry age time gender
## 1 366     1      782 909  127      1
## 2 383     0      806 943  137      1
## 3 402     0      820 957  137      1
## 4 393     0      821 956  135      1
## 5 415     0      823 960  137      1
## 6 396     0      830 940  110      1
```

```
coxfitmen2 <- coxph(Surv(time, death)~ageentry, data =newfitmen2)
summary(coxfitmen2)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ ageentry, data = newfitmen2)
##
```

```
##   n= 56, number of events= 25
##
##              coef exp(coef) se(coef)     z Pr(>|z|)
## ageentry 0.005579  1.005595 0.003271 1.706   0.0881 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## ageentry     1.006     0.9944    0.9992     1.012
##
## Concordance= 0.597  (se = 0.062 )
## Rsquare= 0.05   (max possible= 0.961 )
## Likelihood ratio test= 2.88  on 1 df,   p=0.09
## Wald test            = 2.91  on 1 df,   p=0.09
## Score (logrank) test = 2.96  on 1 df,   p=0.09
```

For time less than 48 months we observe that the hazard ratio for men is between an interval of (.994, 1.008)

We observe for time after 48 months and males, that the hazard ratios lies between the interval (.9992, 1.012).

Part D. We find this is a similar ratio to the group of men who have been at the facility less than 48 months. Since the intervals overlap and contain the possibility of the true hazard ratio being 1, we can conclude there is no differnce between the change in effect of gender, before and after 4 years.