Survival Analysis

Jaeho Hoya Lee

Table of Contents

Setup data

## [1] "/Users/jaehohoyalee/Downloads"

## 'data.frame': 299 obs. of 13 variables:  
## $ age : num 75 55 65 50 65 90 75 60 65 80 ...  
## $ anaemia : int 0 0 0 1 1 1 1 1 0 1 ...  
## $ creatinine\_phosphokinase: int 582 7861 146 111 160 47 246 315 157 123 ...  
## $ diabetes : int 0 0 0 0 1 0 0 1 0 0 ...  
## $ ejection\_fraction : int 20 38 20 20 20 40 15 60 65 35 ...  
## $ high\_blood\_pressure : int 1 0 0 0 0 1 0 0 0 1 ...  
## $ platelets : num 265000 263358 162000 210000 327000 ...  
## $ serum\_creatinine : num 1.9 1.1 1.3 1.9 2.7 2.1 1.2 1.1 1.5 9.4 ...  
## $ serum\_sodium : int 130 136 129 137 116 132 137 131 138 133 ...  
## $ sex : int 1 1 1 1 0 1 1 1 0 1 ...  
## $ smoking : int 0 0 1 0 0 1 0 1 0 1 ...  
## $ time : int 4 6 7 7 8 8 10 10 10 10 ...  
## $ DEATH\_EVENT : int 1 1 1 1 1 1 1 1 1 1 ...

##   
## 0 1   
## 203 96

 Loading Library

## Warning: package 'ggplot2' was built under R version 4.0.2

## Registered S3 methods overwritten by 'tibble':  
## method from   
## format.tbl pillar  
## print.tbl pillar

## Warning: package 'survminer' was built under R version 4.0.2

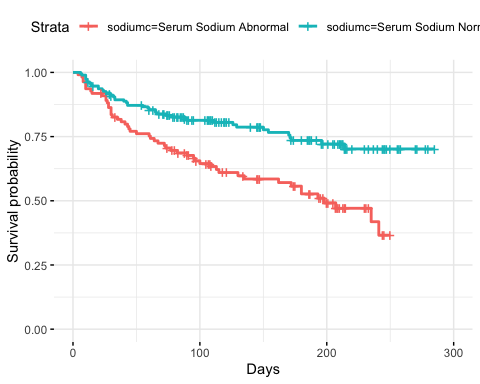
## Loading required package: ggpubr

## Warning: package 'ggpubr' was built under R version 4.0.2

##Kaplan Meier ####-Create categorical variable from continuous variable

heartdata$sodiumc <- ifelse(heartdata$serum\_sodium >135 & heartdata$serum\_sodium<145,  
 "Serum Sodium Normal", "Serum Sodium Abnormal")  
heartdata$efraction <-ifelse(heartdata$ejection\_fraction<=75 & heartdata$ejection\_fraction>=41,   
 "Ejection Normal", "Ejection Abnormal")  
heartdata$sodiumc<- as.factor(heartdata$sodiumc)  
heartdata$efraction<- as.factor(heartdata$efraction)

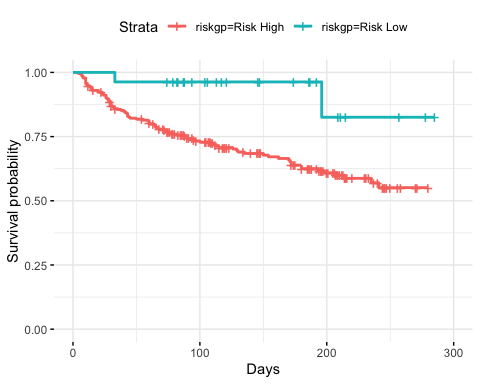
fit\_sd<-survfit(Surv(time,DEATH\_EVENT)~sodiumc, data=heartdata)  
ggsurvplot(fit\_sd,data=heartdata,xlab="Days", ggtheme=theme\_minimal())

 ##Final model for risk factors of Kaplan Meier

heartdata$riskgp <-ifelse(heartdata$age<65  
 & heartdata$efraction=="Ejection Normal"  
 & heartdata$sodiumc=="Serum Sodium Normal", "Risk Low", "Risk High")  
table(heartdata$riskgp)##High: 272, Low: 27

##   
## Risk High Risk Low   
## 272 27

heartdata$riskgp<- as.factor(heartdata$riskgp)  
fit<-survfit(Surv(time,DEATH\_EVENT)~riskgp, data=heartdata)  
ggsurvplot(fit,data=heartdata,xlab="Days", ggtheme=theme\_minimal())

 ##Log rank of P-value for risk group

survdiff(Surv(time, DEATH\_EVENT) ~ riskgp, data=heartdata)

## Call:  
## survdiff(formula = Surv(time, DEATH\_EVENT) ~ riskgp, data = heartdata)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## riskgp=Risk High 272 94 86.27 0.693 6.87  
## riskgp=Risk Low 27 2 9.73 6.145 6.87  
##   
## Chisq= 6.9 on 1 degrees of freedom, p= 0.009

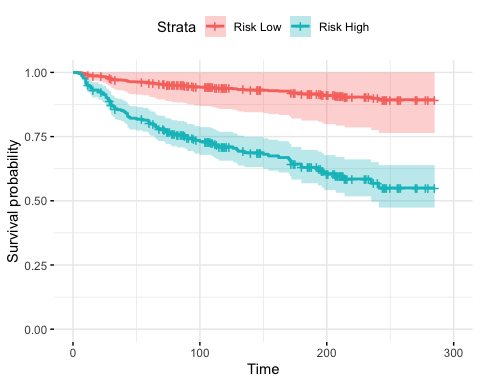
####Difference between group seems significant engough, p= 0.009

###Cox proportional Hazard model

cox.risk<-coxph(Surv(time,DEATH\_EVENT) ~ riskgp, data=heartdata)  
summary(cox.risk)

## Call:  
## coxph(formula = Surv(time, DEATH\_EVENT) ~ riskgp, data = heartdata)  
##   
## n= 299, number of events= 96   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## riskgpRisk Low -1.6708 0.1881 0.7147 -2.338 0.0194 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## riskgpRisk Low 0.1881 5.316 0.04635 0.7633  
##   
## Concordance= 0.543 (se = 0.011 )  
## Likelihood ratio test= 9.84 on 1 df, p=0.002  
## Wald test = 5.47 on 1 df, p=0.02  
## Score (logrank) test = 6.86 on 1 df, p=0.009

Individual\_data=with(heartdata,data.frame(riskgp=c("Risk Low", "Risk High")))  
fit0<- survfit(cox.risk,newdata=Individual\_data)  
riskfactor\_plot=ggsurvplot(fit0,  
 Individual\_data,  
 legend.labs=c("Risk Low","Risk High"),  
 ggtheme = theme\_minimal())  
riskfactor\_plot

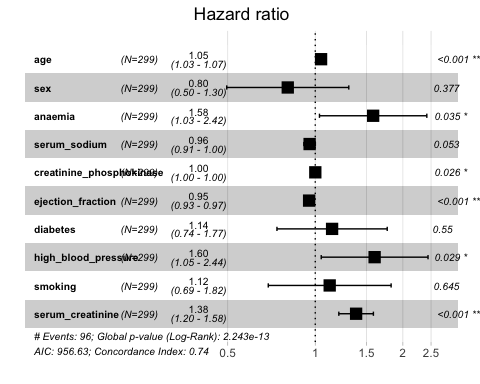
 ##Cox proportional Hazard model

heart.cox<-coxph(Surv(time,DEATH\_EVENT)~age+sex+  
 anaemia+serum\_sodium+creatinine\_phosphokinase+  
 ejection\_fraction+diabetes+high\_blood\_pressure+  
 smoking+serum\_creatinine,data=heartdata)  
summary(heart.cox)

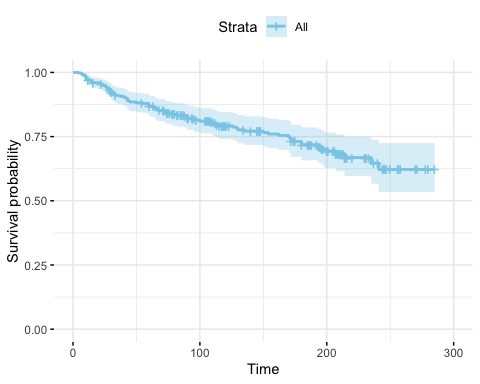
## Call:  
## coxph(formula = Surv(time, DEATH\_EVENT) ~ age + sex + anaemia +   
## serum\_sodium + creatinine\_phosphokinase + ejection\_fraction +   
## diabetes + high\_blood\_pressure + smoking + serum\_creatinine,   
## data = heartdata)  
##   
## n= 299, number of events= 96   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## age 0.0458869 1.0469560 0.0092074 4.984 6.24e-07 \*\*\*  
## sex -0.2177957 0.8042897 0.2464650 -0.884 0.3769   
## anaemia 0.4580205 1.5809414 0.2167654 2.113 0.0346 \*   
## serum\_sodium -0.0448337 0.9561565 0.0231985 -1.933 0.0533 .   
## creatinine\_phosphokinase 0.0002198 1.0002198 0.0000989 2.222 0.0263 \*   
## ejection\_fraction -0.0488823 0.9522932 0.0104769 -4.666 3.08e-06 \*\*\*  
## diabetes 0.1330636 1.1423226 0.2226422 0.598 0.5501   
## high\_blood\_pressure 0.4697306 1.5995632 0.2156587 2.178 0.0294 \*   
## smoking 0.1141273 1.1208948 0.2480591 0.460 0.6455   
## serum\_creatinine 0.3231991 1.3815404 0.0698266 4.629 3.68e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## age 1.0470 0.9551 1.0282 1.0660  
## sex 0.8043 1.2433 0.4962 1.3038  
## anaemia 1.5809 0.6325 1.0337 2.4178  
## serum\_sodium 0.9562 1.0459 0.9137 1.0006  
## creatinine\_phosphokinase 1.0002 0.9998 1.0000 1.0004  
## ejection\_fraction 0.9523 1.0501 0.9329 0.9721  
## diabetes 1.1423 0.8754 0.7384 1.7673  
## high\_blood\_pressure 1.5996 0.6252 1.0482 2.4410  
## smoking 1.1209 0.8921 0.6893 1.8227  
## serum\_creatinine 1.3815 0.7238 1.2048 1.5842  
##   
## Concordance= 0.741 (se = 0.027 )  
## Likelihood ratio test= 81.78 on 10 df, p=2e-13  
## Wald test = 87.61 on 10 df, p=2e-14  
## Score (logrank) test = 88.19 on 10 df, p=1e-14

##Result for hazard ratio visualization of Cox proportional Hazard model

ggforest(heart.cox,data=heartdata)



hf\_plot=ggsurvplot(survfit(heart.cox,data=heartdata),  
 palette = 'skyblue',  
 ggtheme=theme\_minimal())  
hf\_plot

 ##Fitting from our cox-model (high blood pressure relation check)

high\_blood\_pressure\_data=with(heartdata,data.frame(high\_blood\_pressure=c(0,1), anaemia=c(0,0), age=rep(mean(age),2), sex=c(0,0),serum\_sodium=rep(mean(serum\_sodium),2), ejection\_fraction=rep(mean(ejection\_fraction),2),creatinine\_phosphokinase=rep(mean(creatinine\_phosphokinase),2),diabetes=c(0,0),smoking=c(0,0),serum\_creatinine=rep(mean(serum\_creatinine),2)))  
##Strata depending on high\_blood\_pressure (no:0, yes:1)  
fit1=survfit(heart.cox,newdata=high\_blood\_pressure\_data)   
Highbloodpressure\_plot=ggsurvplot(fit1, high\_blood\_pressure\_data, legend.labs=c("High-Bloodpressure1","High\_Bloodpressure2"),ggtheme = theme\_minimal())  
Highbloodpressure\_plot

