PSTAT 175 Project

Group N: Michael Quitana, Robert Hernandez, and Cameron Joe

12/4/2020

Section 1: Descriptions and Covariates

The whas500 data set is provided by Dr. Robert J. Goldberg of the Department of Cardiology at the University of Massachusetts Medical (David W. Hosmer, et al. 13). The goal of the WHAS study is to describe factors associated with trends over time in the incidence and survival rates following hospital admission for acute myocardial infarction (MI). The WHAS study was conducted over a 13 year period starting in 1975 through 2001 with MI patients admitted in hospitals in Massachusetts Standard Metropolitan Statistical Area resulting in more than 11,000 admissions (David W. Hosmer, et al. 13). The whas500 data set is a 23 percent random sample from the years 1997, 1999, and 2001 from the main data of the WHAS study. The data was available on the course Gauchospace, but is publicly accessible on the John Wiley & Sons website ("WHAS Survival").

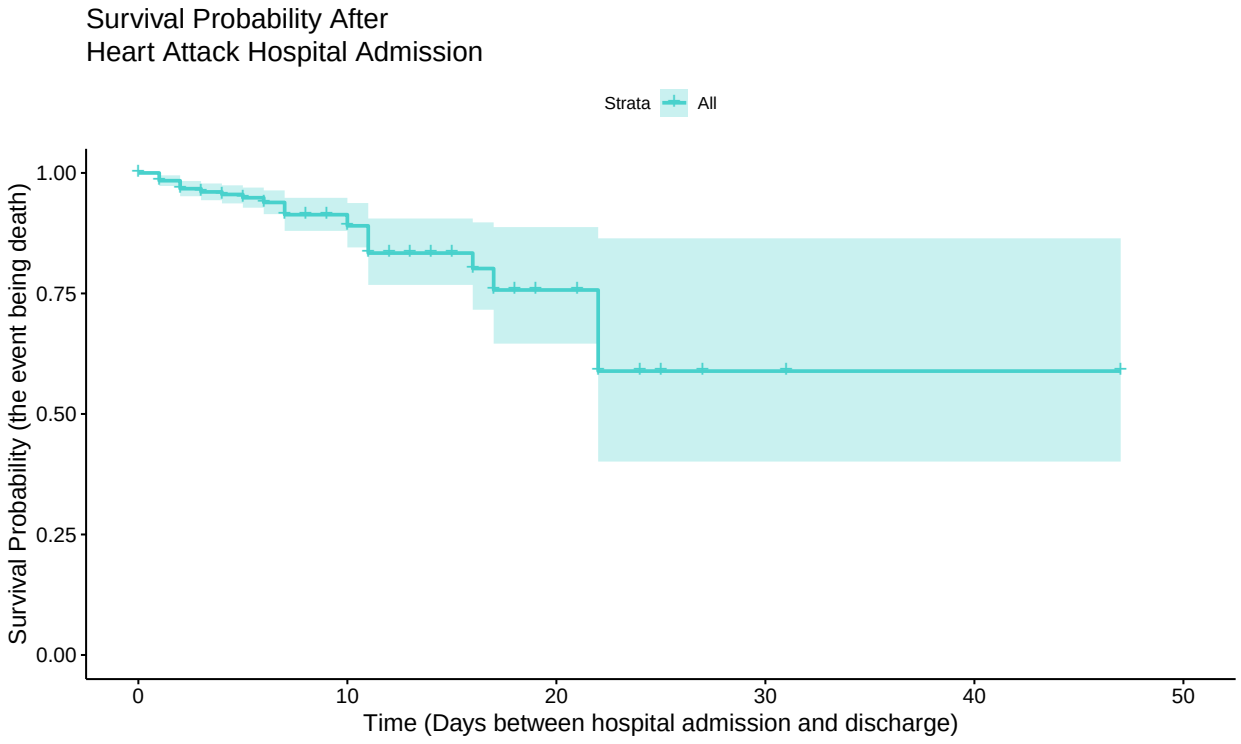
The whas500 dataset consists of 22 variables with the two most important for our survival analysis being los -length of stay at the hospital- and dstat -discharge status (being either 0 for alive or 1 for dead) (David W. Hosmer, et al. 22). There are five continuous covariates to potentially analyze: age (years), initial heart rate (Beats per minute), systolic blood pressure (mmHg), diastolic blood pressure (mmHg), and body mass index (kg/mˆ2). In addition, there are seven binary categorical covariates: history of cardiovascular disease (1 = Yes, 0 = No), atrial fibrillation (Y/N), cardiogenic shock (Y/N), congestive heart complications (Y/N), complete heart block (Y/N), MI order (if it is the patient's first heart attack or not), and MI type (Q-wave or non Q-wave) (David W. Hosmer, et al. 22). There are also categorical variables for gender and cohort year(1997, 1999, 2001). Other variables include the unique patient id and three date variables (for admission, discharge, and last follow-up). Lastly, there is lenfol -the number of days between the last patient follow-up and hospital admission- and fstat -the patient's vital status at this last follow-up with 0 being alive and 1 being dead (David W. Hosmer, et al. 22). The inclusion of these variables allow for the construction of a recurrent events model with two survival analyses.

The primary objective of this paper is to construct a well-fit model of heart attack patient survival rates using the covariates provided in the data. In addition, once a model has been built, we aim to determine how the selected variables affect survival rates (beta estimates and confidence intervals).

In order to accomplish our objectives, we first employ basic Kaplan-Meier plots to better understand our data and covariates. We then use BIC and forward selection to construct a Cox proportional-hazards model that describes the data well but does not overfit. The model's proportional hazards assumption is then checked through a log-log plot and analysis of the residuals. Lastly, a more complex recurrent events model is constructed for the survival rate of the hospital stay and the survival rate after admission.

Section 2: KM Plot

```
library(survival)
library(survminer)
names <- c("id", "age", "gender", "hr", "sysbp", "diasbp", "bmi", "cvd", "afb", "sho", "chf", "av3", "m

whas <- read.table("/Users/anher/Documents/R/R documents/whas500.dat", col.names = names)

whas.fit <- survfit(Surv(los, dstat) ~ 1, data = whas)

ggsurvplot(whas.fit,
     title = "Survival Probability After \nHeart Attack Hospital Admission",
     palette = "mediumturquoise",
     xlab = "Time (Days between hospital admission and discharge)",
     ylab = "Survival Probability (the event being death)"
)
```



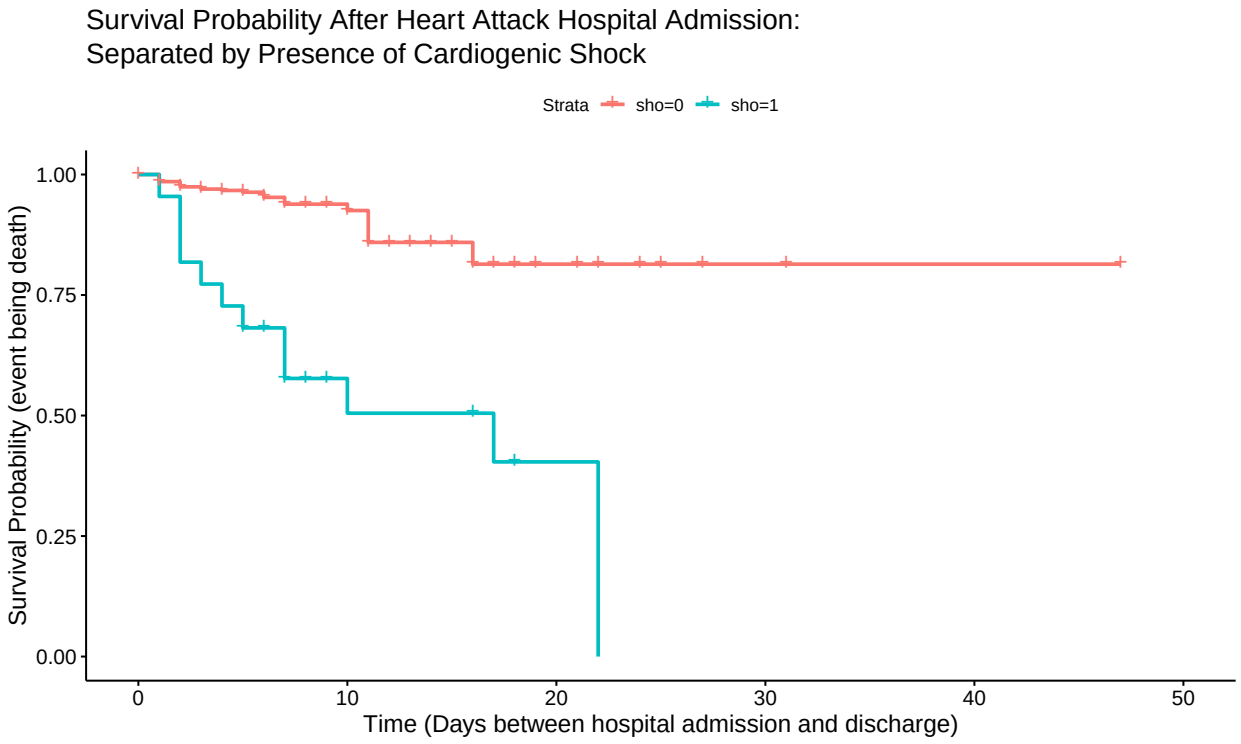
Upon initial glance at the Kaplan-Meier plot, things look to be as we expected. As time goes on, the survival rate continues to decrease. As more patients leave the risk pool by either dying or being censored

(being discharged from the hospital alive), we see that the survival functions begins to move in larger steps for later times.

KM Plot on a few Covariates

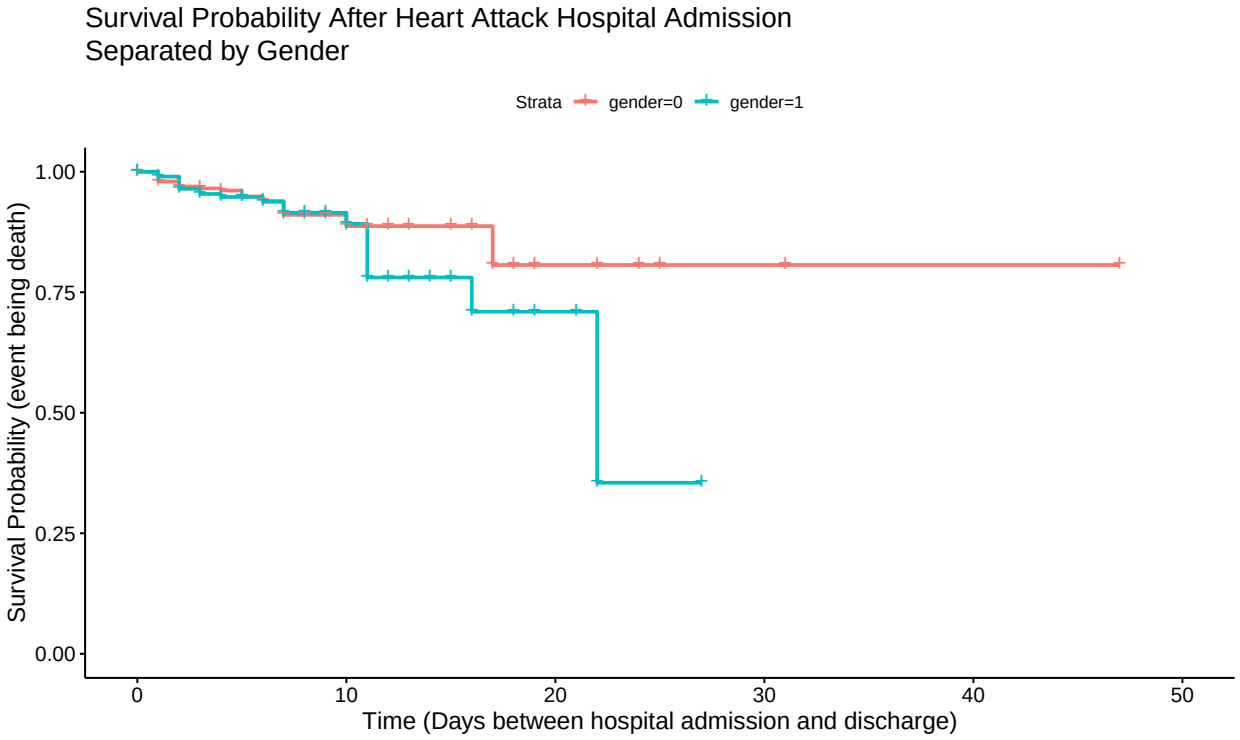
To quickly explore our data, we thought it would be best to construct KM plots for some of the covariates we were interested in to see how they related to survival times. Individual KM plots and brief, rough analyses are provided below.

```
whas.fit.sho <- survfit(Surv(los, dstat) ~ sho, data = whas)

ggsurvplot(whas.fit.sho,
     title = "Survival Probability After Heart Attack Hospital Admission: \nSeparated by Presence of Ca
     xlab = "Time (Days between hospital admission and discharge)",
     ylab = "Survival Probability (event being death)"
)
```



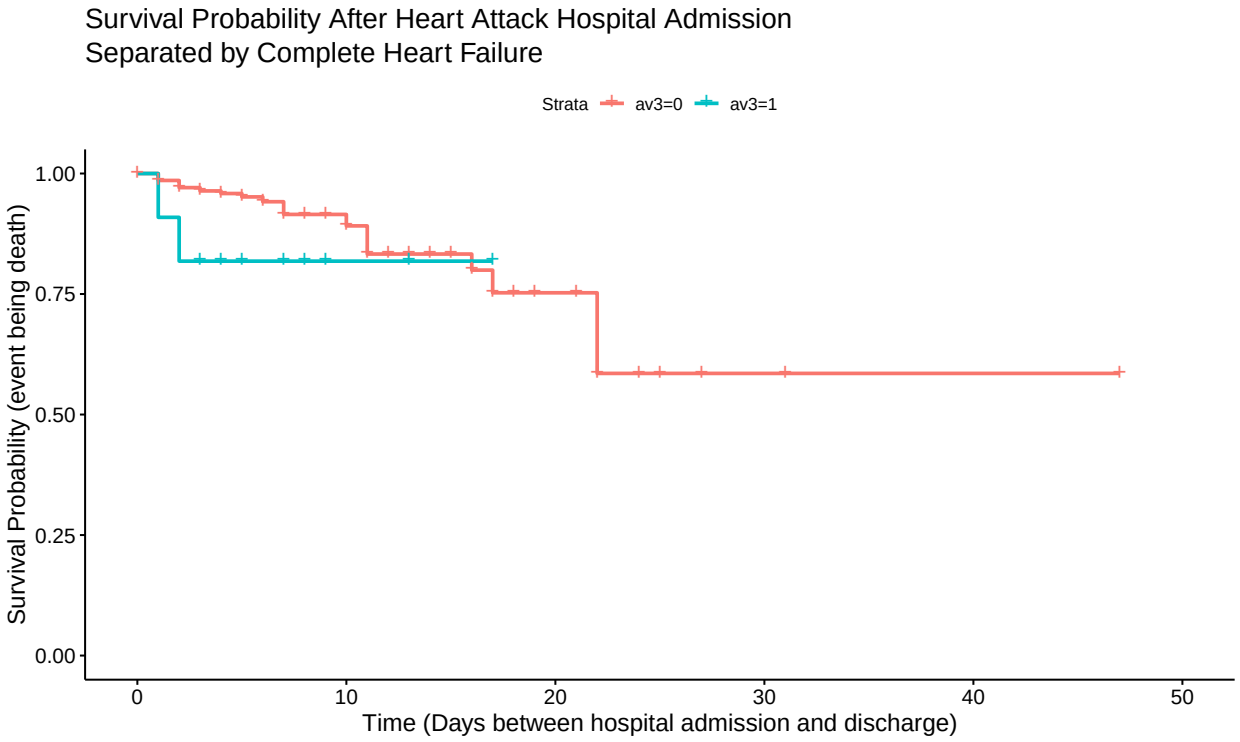
The two plots deviate greatly from each other, implying that the presence of cardiogenic shock greatly reduces survival probabilities.

```
whas.fit.gender <- survfit(Surv(los, dstat) ~ gender, data = whas)

ggsurvplot(whas.fit.gender,
     title = "Survival Probability After Heart Attack Hospital Admission \nSeparated by Gender",
     xlab = "Time (Days between hospital admission and discharge)",
     ylab = "Survival Probability (event being death)"
)
```



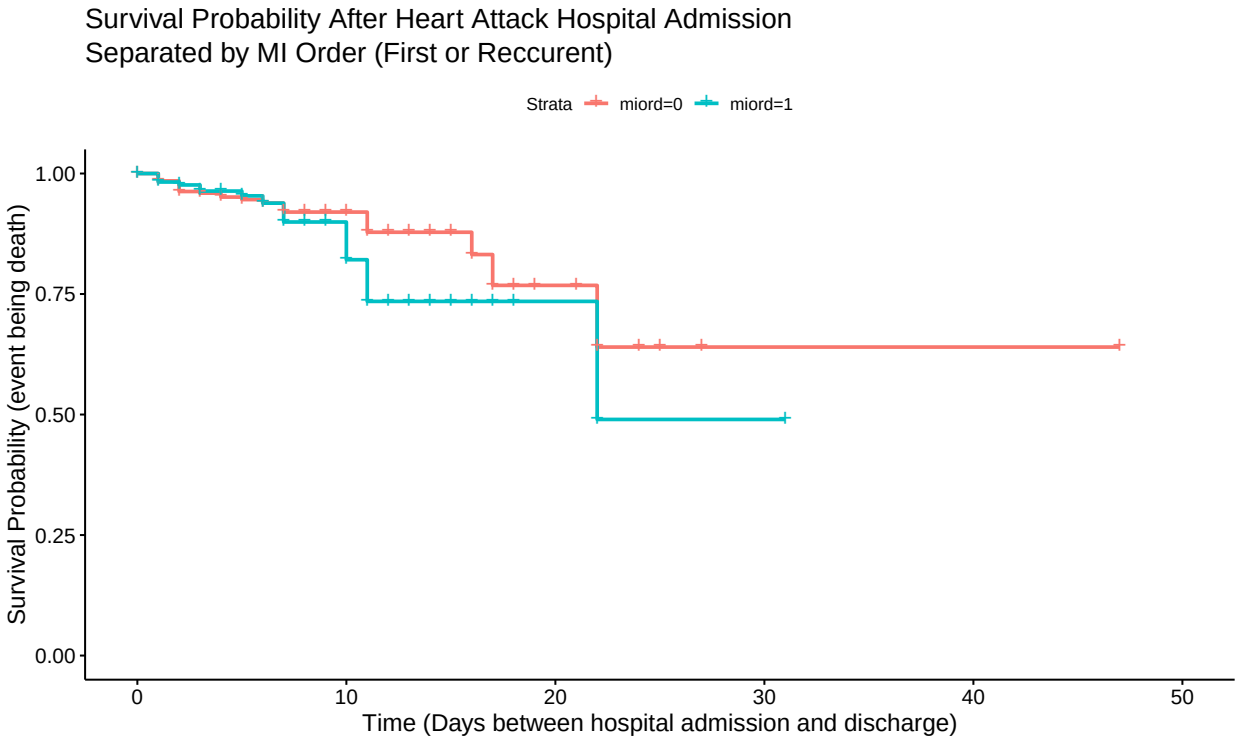
The difference between genders is less apparent than cardiogenic shock (especially within the 10 day range). But, it looks as if males (marked as 1) have a slightly lower survival probability in comparison to females.

```
whas.fit.av3 <- survfit(Surv(los, dstat) ~ av3, data = whas)

ggsurvplot(whas.fit.av3,
     title = "Survival Probability After Heart Attack Hospital Admission \nSeparated by Complete Heart
     xlab = "Time (Days between hospital admission and discharge)",
     ylab = "Survival Probability (event being death)"
)
```



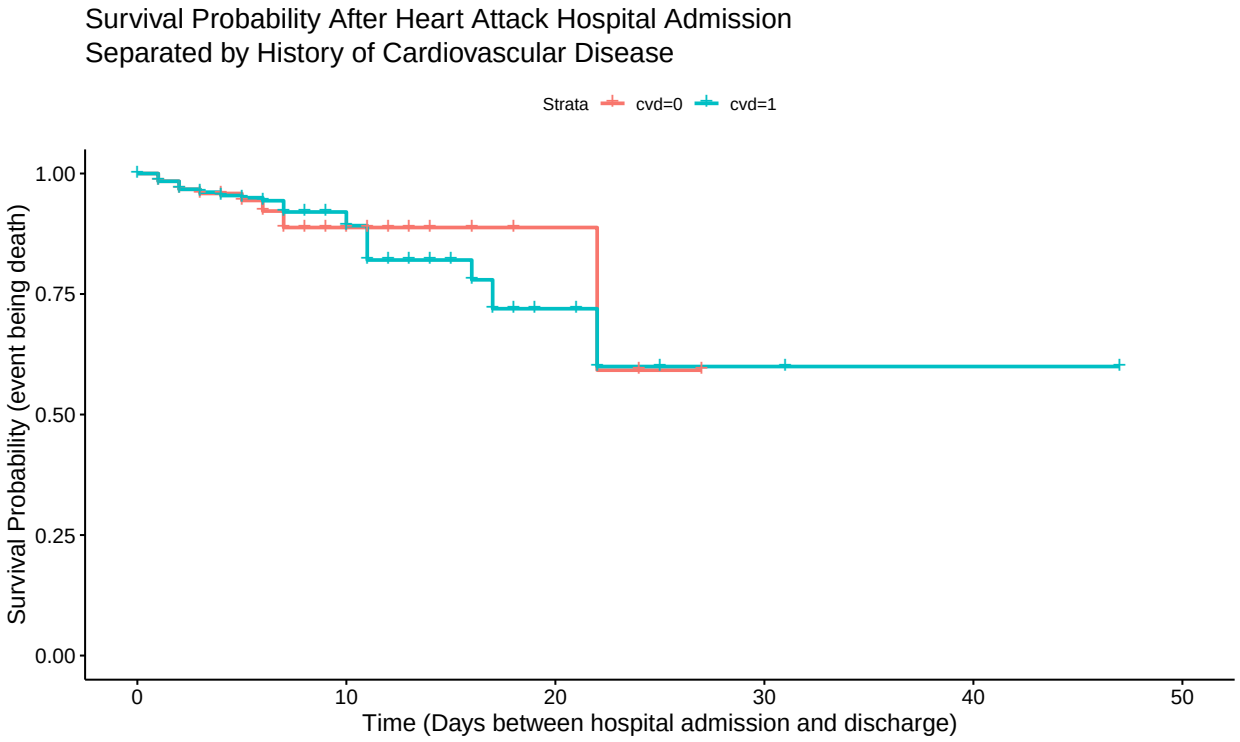
Looking at the graph, it appears that those that experience complete heart failure have a lower survival probability. This cannot be said confidently as there is a smaller sample size for those who have experienced complete heart failure.

```
whas.fit.miord <- survfit(Surv(los, dstat) ~ miord, data = whas)

ggsurvplot(whas.fit.miord,
     title = "Survival Probability After Heart Attack Hospital Admission \nSeparated by MI Order (First
     xlab = "Time (Days between hospital admission and discharge)",
     ylab = "Survival Probability (event being death)"
)
```



Looking at the graph, it appears that those who have already had a prior heart attack have a lower survival probability than those who are experiencing their first heart attack.

```
whas.fit.cvd <- survfit(Surv(los, dstat) ~ cvd, data = whas)

ggsurvplot(whas.fit.cvd,
     title = "Survival Probability After Heart Attack Hospital Admission \nSeparated by History of Cardi
     xlab = "Time (Days between hospital admission and discharge)",
     ylab = "Survival Probability (event being death)"
)
```



When separating on history of cardiovascular disease, it is difficult to determine the relationship between the two groups. This covariate -along with all other prior KM plot analyses- will be explored at more depth in further sections.

Section 3: Coxph Model fitting

First Round of Fitting

```
whas1 <- coxph(Surv(los, dstat) ~ age, data = whas)
whas2 <- coxph(Surv(los, dstat) ~ gender, data = whas)
whas3 <- coxph(Surv(los, dstat) ~ hr, data = whas)
whas4 <- coxph(Surv(los, dstat) ~ sysbp, data = whas)
whas5 <- coxph(Surv(los, dstat) ~ diasbp, data = whas)
whas6 <- coxph(Surv(los, dstat) ~ bmi, data = whas)
whas7 <- coxph(Surv(los, dstat) ~ cvd, data = whas)
whas8 <- coxph(Surv(los, dstat) ~ afb, data = whas)
whas9 <- coxph(Surv(los, dstat) ~ sho, data = whas)
whas10 <- coxph(Surv(los, dstat) ~ chf, data = whas)
whas11 <- coxph(Surv(los, dstat) ~ av3, data = whas)
whas12 <- coxph(Surv(los, dstat) ~ miord, data = whas)
whas13 <- coxph(Surv(los, dstat) ~ mitype, data = whas)
whas14 <- coxph(Surv(los, dstat) ~ year, data = whas)

BIC(whas1, whas2, whas3, whas4, whas5, whas6, whas7, whas8, whas9, whas10, whas11, whas12, whas13, whas
```

```
##        df      BIC
```

```
## whas1   1 397.0925
## whas2   1 414.6859
## whas3   1 415.5095
## whas4   1 401.1719
## whas5   1 406.1335
## whas6   1 409.9470
## whas7   1 416.1201
## whas8   1 416.1258
## whas9   1 393.7014
## whas10  1 410.5048
## whas11  1 415.1969
## whas12  1 415.5882
## whas13  1 415.8470
## whas14  1 415.1381
```

Using forward selection to create the model, for the first round of model fitting, we created coxph models using each covariate. We chose to use BIC for model evaluation as it punishes the inclusion of additional covariates harder than AIC and we would like to limit the number of covariates in the model to avoid overfitting. For the first round of models, the model using sho or cardiogenic shock was found to yield the lowest BIC of 393.7014.

Successive rounds of fitting

```
whas9.1 <- coxph(Surv(los, dstat) ~ sho + age, data = whas)
whas9.1.3 <- coxph(Surv(los, dstat) ~ sho + age + sysbp, data = whas)
whas9.1.3.2 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + hr, data = whas)

BIC(whas9.1, whas9.1.3, whas9.1.3.2)
```

```
##             df      BIC
## whas9.1      2 382.6642
## whas9.1.3    3 378.7440
## whas9.1.3.2  4 380.9133
```

For following rounds of model fitting, each of the remaining covariates were added to the previous model and tested for BIC (code for all iterations included in the appendix). The lowest BIC model would be kept and built upon in following iterations. In the second round of model fitting, it was found that the model using both sho + age was found to yield the lowest BIC of 382.6642. For the third round of model fitting, it was found that the model using sho + age + sysbp (or systolic blood pressure) was found to yield the lowest BIC of 378.7440. For the fourth round of model fitting, it was found that the model using sho + age + sysbp + hr (heart rate) was found to yield the lowest BIC of 380.9133. As this model yielded a higher BIC than the last iteration, we know that the third model is the best for modeling survival rates and including any other covariates would lower the BIC.

Finding coefficient estimates and confidence intervals

```
# P-values
summary(whas9.1.3)$coefficients[, 5]
```

```
##          sho          age        sysbp
## 3.815937e-05 5.590685e-04 9.884368e-03
```

```
# Estimates
exp(coef(whas9.1.3))
```

```
##       sho       age     sysbp
## 4.4478907 1.0571523 0.9862268
```

```
# 95% CI
exp(confint(whas9.1.3))
```

```
##           2.5 %    97.5 %
## sho   2.1862062 9.0493437
## age   1.0243014 1.0910567
## sysbp 0.9758899 0.9966731
```

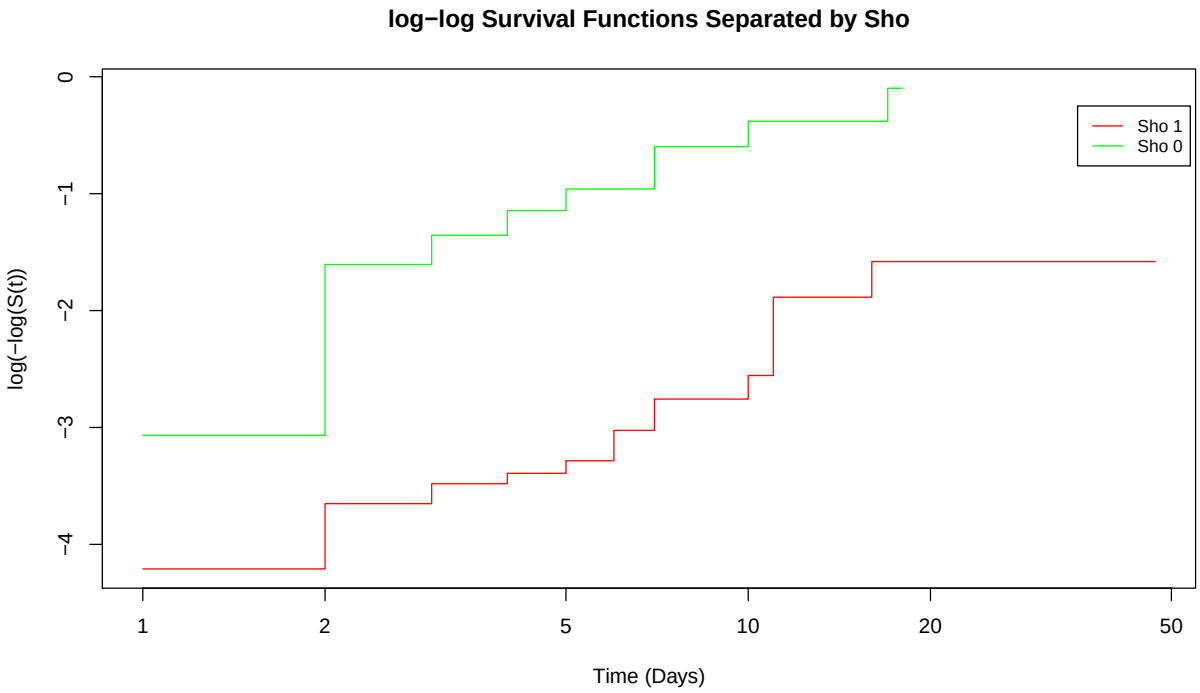
```
# Likelihood ratio p-value
summary(whas9.1.3)
```

```
## Call:
## coxph(formula = Surv(los, dstat) ~ sho + age + sysbp, data = whas)
##
##   n= 500, number of events= 39
##
##            coef exp(coef)  se(coef)      z Pr(>|z|)
## sho    1.492430  4.447891  0.362385  4.118 3.82e-05 ***
## age    0.055579  1.057152  0.016106  3.451 0.000559 ***
## sysbp -0.013869  0.986227  0.005376 -2.580 0.009884 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##       exp(coef) exp(-coef) lower .95 upper .95
## sho      4.4479     0.2248    2.1862    9.0493
## age      1.0572     0.9459    1.0243    1.0911
## sysbp    0.9862     1.0140    0.9759    0.9967
##
## Concordance= 0.789  (se = 0.04 )
## Likelihood ratio test= 44.71  on 3 df,   p=1e-09
## Wald test            = 45.1  on 3 df,   p=9e-10
## Score (logrank) test = 56.48  on 3 df,   p=3e-12
```

The sho covariate was found to be significant with a p-value of 3.82e-05, a β estimate of 4.447891, and a standard 95% confidence interval of (2.1862062, 9.0493437). So, those with cardiogenic shock have a higher hazard rate (with the event being death) in comparison to those who did not experience cardiogenic shock. The age covariate was found to be significant with a p-value of 0.000559, a β estimate of 1.057152, and a confidence invterval of (1.0243014, 1.0910567). So, a one year age increase yields a slightly higher hazard rate. The sysbp covariate was found to be significant with a p-value of 0.009884, a β estimate of 0.986227, and a a confidence interval of (0.9758899, 0.9966731). So, an increase in systolic blood pressure is associated with a slightly lower hazard rate (death rate). A low p-value of 1e-09 for the likelihood ratio test indicates the significance of the overall model.

Part 4: Checking Coxph Assumptions

Observation of Log-log plot



The log-log plot hints that the proportional hazards assumption has not been violated as the two functions do not cross and are close to being parallel. Since the proportional hazards is not violated, the ratio of the hazards is constant over time.

Cox.zph testing

```
whas9.1.3.zph <- cox.zph(whas9.1.3, transform="rank")
whas9.1.3.zph
```

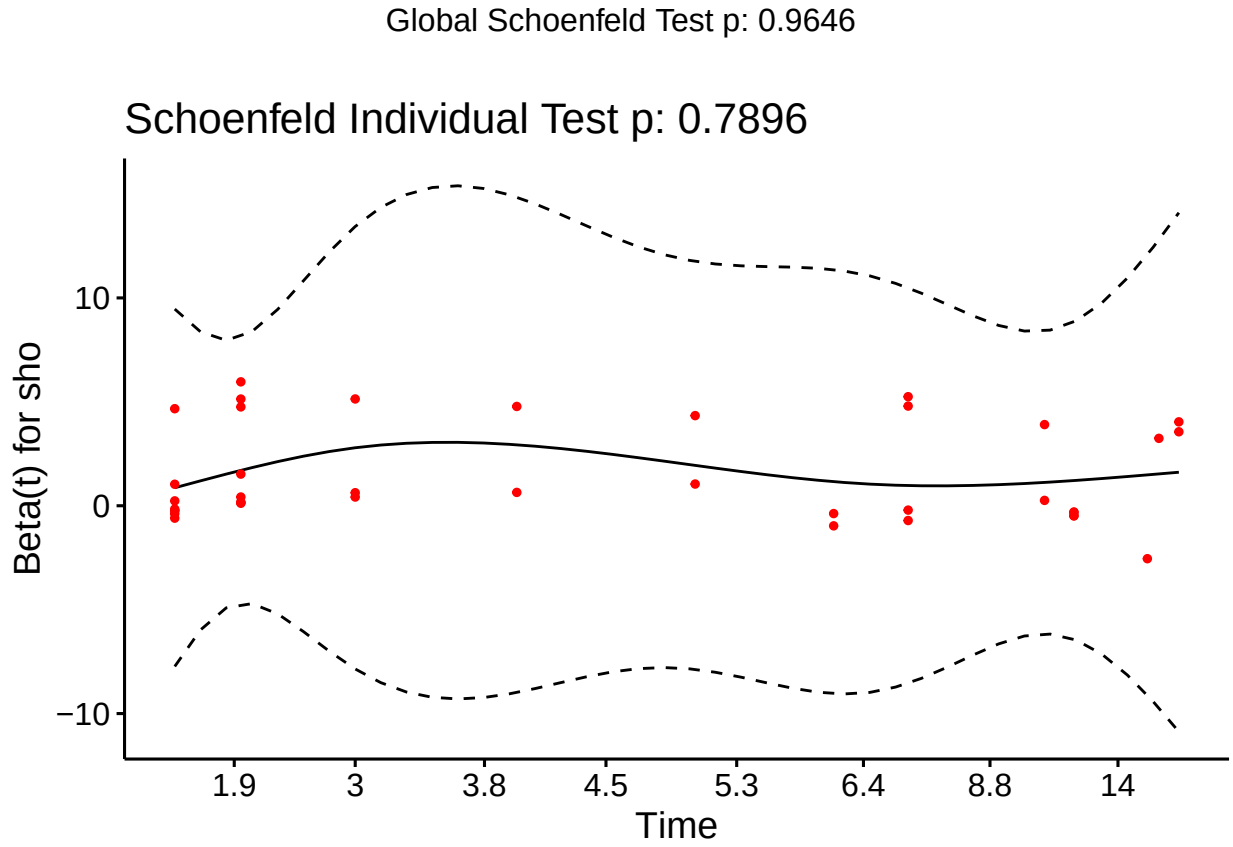
```
##         chisq df    p
## sho    0.0712  1 0.79
## age    0.0929  1 0.76
## sysbp  0.1240  1 0.72
## GLOBAL 0.2751  3 0.96
```

All of the covariates have a large p-value, so we fail to reject the null hypothesis that the model follows proportional hazards assumptions. This is good for our model as the model continues to satisfy the proportional hazards assumption. So, we will not need to use time varying coefficients to make the model proportional.

Residuals Plots

```
whas9.1.3.test <- cox.zph(
  coxph(Surv(los,dstat)~sho, data=whas), transform="rank" )
whas9.1.3.test
```

```
##         chisq df    p
## sho     0.417  1 0.52
## GLOBAL 0.417  1 0.52
```



The Schoenfield test plot shows that the residuals closely resemble horizontal lines, so our model still satisfies the proportional hazards assumption. Also, applying coxzph to the sho covariate results in a large p-value of 0.52, so once again we fail to reject the null hypothesis that the model follows the proportional hazards assumption.

Part 5 Recurrent Events

We chose to use recurrent event models in our project as they are able to represent the differences between time in the hospital, and time outside of the hospital.

We will be looking at the covariates, Cardiogenic Shock(sho), Age in Years(age), and Initial Systolic Blood Pressure(sysbp) as we found these to be the most significant variables in our model fitting process.

First Gap(Marginal) Model

For our first gap model, we are looking at length of stay in the hospital and status upon release.

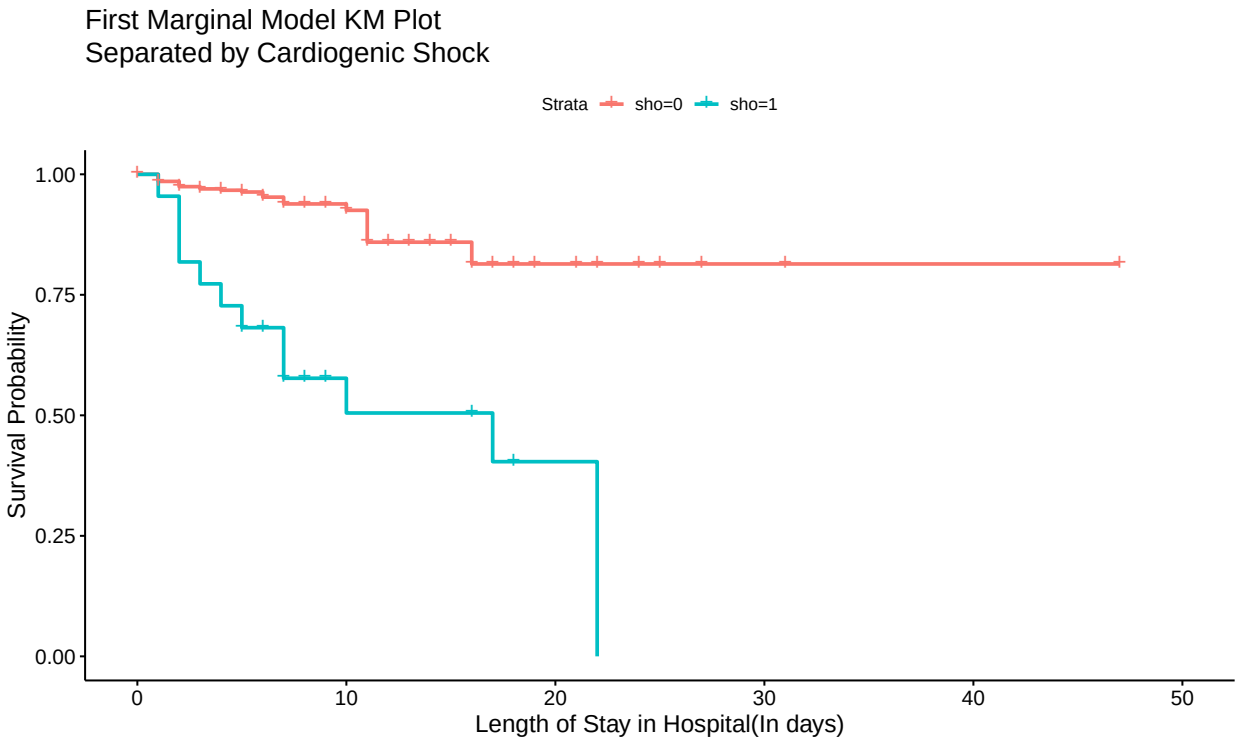
From our KM graph displaying Cardiogenic Shock, we see patients suffering from this illness have a much lower survival probability than patients without shock.

The age box plots displays that older patients seem to die more frequently than young patients. While the Systolic Blood pressure box plots has a more even distribution

Our CoxPH and confidence intervals agree with our graphs for the most part. Cardiogenic Shock has a hazard rate of 4.44 and a confidence interval above 2.[**2.18620,9.04934**] As well as a small p-value proving it's significance.

Age is also significant with a small p-value, a 1 year increase in age leads to around a 6% increase in hazard rate. The confidence interval also does not contain 1. Systolic Blood Pressure is also significant, a 1 unit increase in blood pressure leads to a 1.5% drop in hazard rate. The p-value displays this is significant and our confidence interval not containing one confirms that.

```
marg.surv1 <- Surv(whas$los, whas$dstat)
marg.fit1 <- surv_fit(marg.surv1 ~ sho, data = whas)
par(mfrow=c(2,2))
ggsurvplot(marg.fit1, title = "First Marginal Model KM Plot \nSeparated by Cardiogenic Shock", ylab="Su
```




```
boxplot(whas$age~whas$dstat,main = "Boxplot of Initial Age (Years)",xlab="Status(0=Alive)(1=Dead) of pa
par(mar=c(5,5,4,1)+.1)
boxplot(whas$sysbp~whas$dstat,main = "Boxplot of Systolic Blood Pressure",xlab="Status(0=Alive)(1=Dead)

#mean(whas$sysbp) #144.7
#mean(whas$age) # 69.8
```

```
#coxph and CI
coxph(marg.surv1 ~ sho+age+sysbp, data = whas)
```

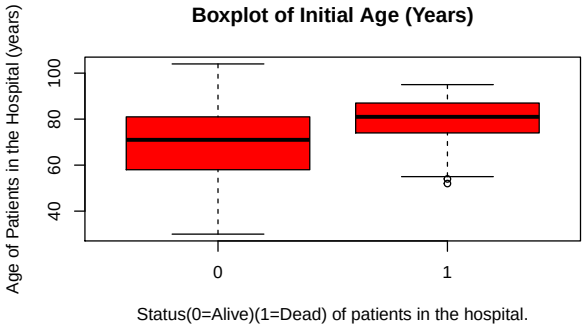
```
## Call:
## coxph(formula = marg.surv1 ~ sho + age + sysbp, data = whas)
##
##             coef exp(coef)  se(coef)      z        p
## sho    1.492430  4.447891  0.362385  4.118 3.82e-05
## age    0.055579  1.057152  0.016106  3.451 0.000559
## sysbp -0.013869  0.986227  0.005376 -2.580 0.009884
##
## Likelihood ratio test=44.71  on 3 df, p=1.065e-09
## n= 500, number of events= 39
```

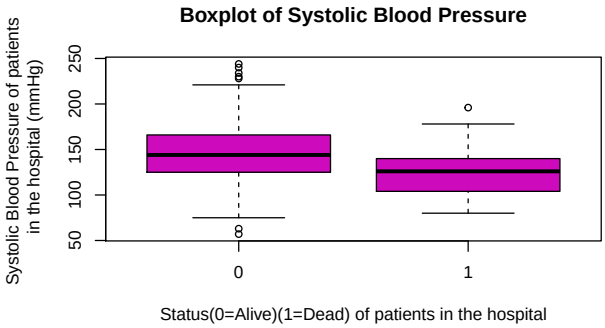
```
CI1 <- coxph(marg.surv1 ~ sho+age+sysbp, data = whas)
exp(coef(CI1))
```

```
##       sho       age     sysbp
## 4.4478907 1.0571523 0.9862268
```

```
as.table(exp(confint(CI1)))
```

```
##           2.5 %    97.5 %
## sho   2.1862062 9.0493437
## age   1.0243014 1.0910567
## sysbp 0.9758899 0.9966731
```





Second Gap Model Adjustments(Trials and Tribulations)

Here we have our Cox Proportional hazard rates for our second gap model measuring the covariates but now for time outside of the hospital. Before we can start analyzing these hazard rates there is an adjustment we need to make in our model. As you can see from our 6th observation below, subjects who died in the hospital(dstat=1) are not factored out of our gap model for subjects outside the hospital(fstat=1).

In order to avoid overestimating our hazard rate we will need to make an adjustment to our data. We will not be using data from this Cox PH model.

```
marg.surv3 <- Surv(whas$lenfol-whas$los,whas$fstat)
coxph(marg.surv3~whas$sho+whas$age+whas$sysbp)
```

```
## Call:
## coxph(formula = marg.surv3 ~ whas$sho + whas$age + whas$sysbp)
##
##                 coef exp(coef)  se(coef)      z        p
## whas$sho    0.998535  2.714303  0.269911  3.699 0.000216
## whas$age    0.065729  1.067937  0.006111 10.756  < 2e-16
## whas$sysbp -0.003101  0.996904  0.002117 -1.465 0.143002
##
## Likelihood ratio test=154.9  on 3 df, p=< 2.2e-16
## n= 500, number of events= 215
```

```
whas[6,]
```

```
##   id age gender hr sysbp diasbp      bmi cvd afb sho chf av3 miord mitype year
## 6  6  70      0 76    83     54 23.24236   1   0   0   0   1     0      0    1
##    admitdate    disdate      fdate los dstat lenfol fstat
## 6 03/11/1997 03/12/1997 03/12/1997   1     1      1     1
```

```
whas$dstat[whas$id==6]
```

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

```
whas$fstat[whas$id==6]
```

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

Adjusted Second Gap Model(Time outside of the hospital)

The adjustment we made was removing the 39 observations where subjects died in the hospital, now we are only viewing the 461 subject who made it out of the hospital.

```
sum(whas$dstat==1)
```

```
## [1] 39
```

```
subset.whas <- whas[which(whas$dstat==0), ]
```

Graphs For Second Gap Model

These graphs give us an idea of what might have happened during the length of time outside of the hospital, which we can see is well over 100 times greater than our length of stay inside the hospital. For cardiogenic shock KM Plot we see that the two lines intersect multiple times, which means these survival probabilities are maybe similar.

From our age box plots it seems that older patients are more likely to die, while younger patients stay alive. From our systolic blood pressure box plots it seems there is a even distribution between the two status, indicating they may be similar.

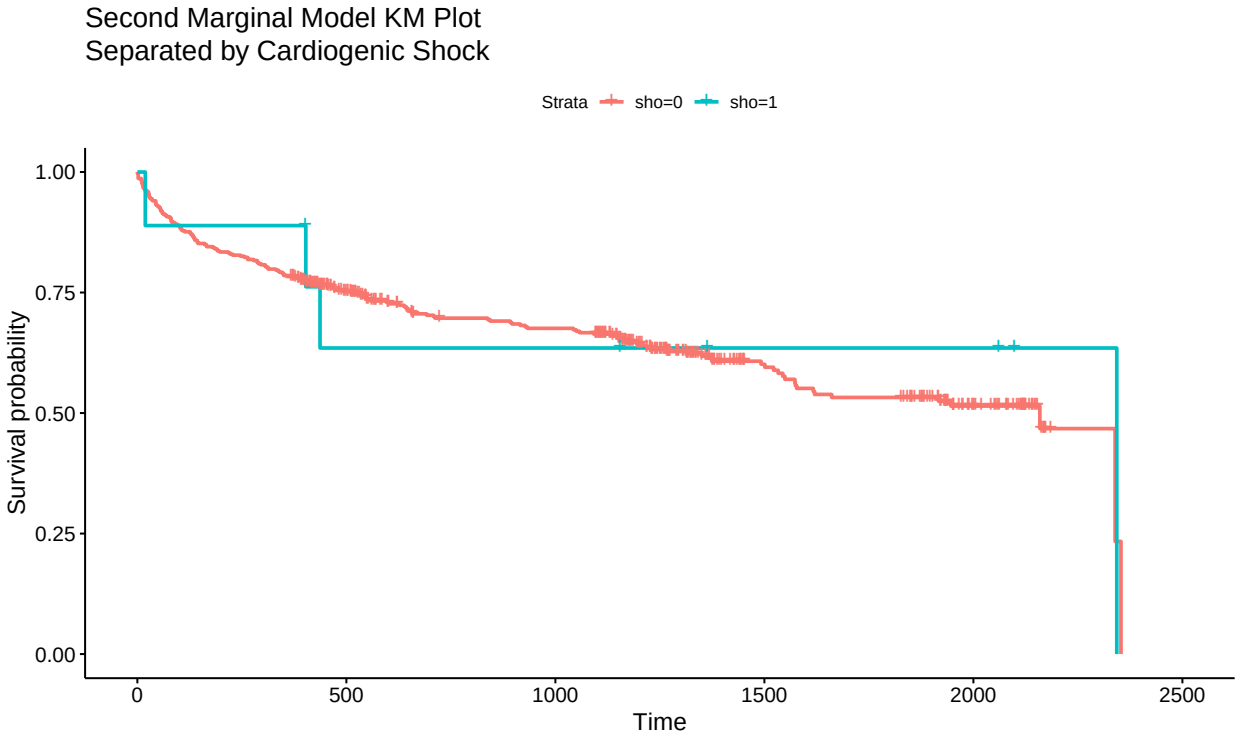
```
mean(subset.whas$los)
```

```
## [1] 6.130152
```

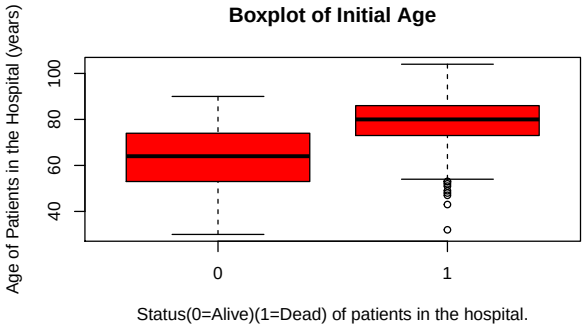
```
mean(subset.whas$lenfol)
```

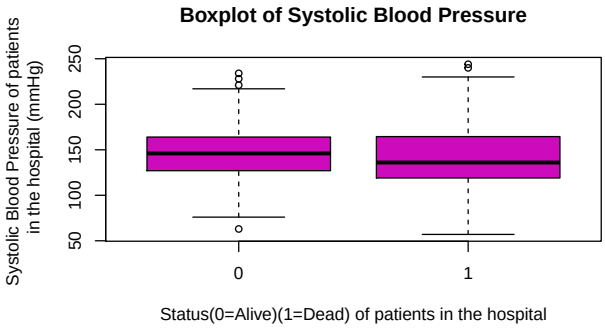
```
## [1] 956.5857
```

```
marg.surv3 <- Surv(subset.whas$lenfol-subset.whas$los, subset.whas$fstat)
marg.fit3 <- surv_fit(marg.surv3 ~ sho, data = subset.whas)
ggsurvplot(title = "Second Marginal Model KM Plot \nSeparated by Cardiogenic Shock", marg.fit3)
```



```r
par(mfrow=c(2,2))

boxplot(whas$age~whas$fstat,main = "Boxplot of Initial Age",xlab="Status(0=Alive)(1=Dead) of patients i
par(mar=c(5,5,4,1)+.1)
boxplot(whas$sysbp~whas$fstat,main = "Boxplot of Systolic Blood Pressure",xlab="Status(0=Alive)(1=Dead)
```





Cox PH and CI for our Second Gap Model

From our CoxPH for our second gap model displaying time outside the hospital, we see one of these covaraites behaves very differently than our first gap model. Patients with Cardiogenic Shock actually have lower hazard rates(68% of Non-Shock Patients), however this isn't statistically significant as the p-value for shock is large at 0.493. We believe this great change in hazard rates is due to the small amount of patients who had Cardiogenic Shock(9 Subjects out of 461).**CI[0.22632,2.04588]** Is very large and contains 1, displaying the large amount of variation in this value and further displaying the insignificance of the coefficient.

Our Age covariate values are slightly increased in comparison to the first gap model. We see that as subjects age a year, their hazard rate increases by 6%. This is statistically significant as our p-value is very small. **CI[1.05599,1.08401]** Doesn't contain 1, further showing that these differences between ages are significant.

Our systolic blood pressure covariate values also stay similar to the first gap model. We see that as subjects blood pressure increases by one, their hazard ratio actually ever so slightly decreases. However, our p-value is large, displaying that these differences are not statistically significant unlike the first model, where Systolic Blood Pressure had a hazard ratio less than 1 and was significant. **CI[0.99426,1.00362]** Contains 1, further showing the insignificance of coefficient.

```
coxph(marg.surv3~sho+age+sysbp,data=subset.whas)
```

```
## Call:
## coxph(formula = marg.surv3 ~ sho + age + sysbp, data = subset.whas)
##
##            coef exp(coef)  se(coef)      z      p
## sho   -0.384974  0.680468  0.561646 -0.685  0.493
## age    0.067574  1.069910  0.006682 10.113 <2e-16
## sysbp -0.001076  0.998925  0.002391 -0.450  0.653
##
## Likelihood ratio test=121.7  on 3 df, p=< 2.2e-16
## n= 461, number of events= 176
```

```
CI3 <-coxph(marg.surv3~sho+age+sysbp,data=subset.whas)
exp(coef(CI3))
```

```
##       sho       age     sysbp
## 0.6804684 1.0699099 0.9989249
```

```
as.table(exp(confint(CI3)))
```

```
##           2.5 %    97.5 %
## sho   0.2263258 2.0458880
## age   1.0559888 1.0840145
## sysbp 0.9942556 1.0036162
```

```
sum(subset.whas$sho==1)
```

```
## [1] 9
```

Cox PH Without Shock

In our first gap model, many patients died of Cardiogenic Shock, as a result there were only 9 subjects with Cardiogenic Shock in our second gap model. We chose to remove Shock in our First and Second Gap Model, so it will not bias our Systolic Blood Pressure and Age Covaraites.

For our First Gap Model, nothing really changes as we remove Cardiogenic Shock, the coefficients are the same but the p-values for age and systolic blood pressure are slightly smaller. The confidence intervals are also very similar to the first gap model

For our Second Gap Model, again nothing really changes as we remove Cardiogenic Shock. Age stays the same, and Systolic Blood pressures coefficient stays the same, but the p-value actually increases, making it more insignificant than before. The confidence intervals are almost equivalent to the origial second gap model.

```
#Gap 1
coxph(marg.surv1~age+sysbp,data=whas)
```

```
## Call:
## coxph(formula = marg.surv1 ~ age + sysbp, data = whas)
```

```
##
##           coef exp(coef) se(coef)      z        p
## age    0.05630   1.05791  0.01567  3.593 0.000327
## sysbp -0.01802   0.98214  0.00559 -3.225 0.001261
##
## Likelihood ratio test=30.5  on 2 df, p=2.378e-07
## n= 500, number of events= 39
```

```
#Gap 2
coxph(marg.surv3~age+sysbp,data=subset.whas)
```

```
## Call:
## coxph(formula = marg.surv3 ~ age + sysbp, data = subset.whas)
##
##              coef  exp(coef)   se(coef)      z      p
## age    0.0675742  1.0699097  0.0066973 10.090 <2e-16
## sysbp -0.0008974  0.9991030  0.0023576 -0.381  0.703
##
## Likelihood ratio test=121.2  on 2 df, p=< 2.2e-16
## n= 461, number of events= 176
```

```
as.table(exp(confint(coxph(marg.surv1~age+sysbp,data=whas))))
```

```
##           2.5 %    97.5 %
## age   1.0259203 1.0909040
## sysbp 0.9714351 0.9929553
```

```
as.table(exp(confint(coxph(marg.surv3~age+sysbp,data=subset.whas))))
```

```
##           2.5 %    97.5 %
## age   1.0559574 1.0840463
## sysbp 0.9944971 1.0037302
```

Section 6: Conclusion

Through forward selection with BIC as a source of model evaluation, it was found that the best model include cardiogenic shock, age, and systolic blood pressure. In specific, presence of cardiogenic shock yield significantly higher likelihoods of death (~4.5 times) in comparison to those not having cardiogenic shock. The proportional hazards assumptions was satisfied by multiple tests such as observing the Shoenfield residuals, cox.zph testing, and log-log plot to confirm confidently of our model.

The Recurrent Event Models gave us a new way to view our covariates. We were able to see the great difference in Cardiogenic Shock's hazard rate when comparing the two gap models. Which showed Cardiogenic Shock was significant overall in increasing the hazard rate, just less than we would have thought. The gap models also displayed that an increase in age also resulted in an increase in hazard rate. Lastly we saw that Systolic blood pressure isn't a very significant covariate. We made a model without Cardiogenic Shock(as it was such a strong predictor) to see if it had an effect on the other covariates. The effect it ended up having was very miniscule.

The most important thing that our recurrent events model was able to show was that Cardiogenic Shock has a very high hazard rate initially, but after that initial period the hazard rate for shock patients is either insignificantly different or slightly lower than non-shock patients. The small sample size of Cardiogenic Shock patients throws noise into these numbers, but a very interesting result from these models.

Appendix (Model Fitting)

First Round of Fitting

```
whas1 <- coxph(Surv(los, dstat) ~ age, data = whas)
whas2 <- coxph(Surv(los, dstat) ~ gender, data = whas)
whas3 <- coxph(Surv(los, dstat) ~ hr, data = whas)
whas4 <- coxph(Surv(los, dstat) ~ sysbp, data = whas)
whas5 <- coxph(Surv(los, dstat) ~ diasbp, data = whas)
whas6 <- coxph(Surv(los, dstat) ~ bmi, data = whas)
whas7 <- coxph(Surv(los, dstat) ~ cvd, data = whas)
whas8 <- coxph(Surv(los, dstat) ~ afb, data = whas)
whas9 <- coxph(Surv(los, dstat) ~ sho, data = whas)
whas10 <- coxph(Surv(los, dstat) ~ chf, data = whas)
whas11 <- coxph(Surv(los, dstat) ~ av3, data = whas)
whas12 <- coxph(Surv(los, dstat) ~ miord, data = whas)
whas13 <- coxph(Surv(los, dstat) ~ mitype, data = whas)
whas14 <- coxph(Surv(los, dstat) ~ year, data = whas)

BIC(whas1, whas2, whas3, whas4, whas5, whas6, whas7, whas8, whas9, whas10, whas11, whas12, whas13, whas
```

```
##        df      BIC
## whas1   1 397.0925
## whas2   1 414.6859
## whas3   1 415.5095
## whas4   1 401.1719
## whas5   1 406.1335
## whas6   1 409.9470
## whas7   1 416.1201
## whas8   1 416.1258
## whas9   1 393.7014
## whas10  1 410.5048
## whas11  1 415.1969
## whas12  1 415.5882
## whas13  1 415.8470
## whas14  1 415.1381
```

Second Round of Fitting

```
whas9.1 <- coxph(Surv(los, dstat) ~ sho + age, data = whas)
whas9.2 <- coxph(Surv(los, dstat) ~ sho + gender, data = whas)
whas9.3 <- coxph(Surv(los, dstat) ~ sho + hr, data = whas)
whas9.4 <- coxph(Surv(los, dstat) ~ sho + sysbp, data = whas)
whas9.5 <- coxph(Surv(los, dstat) ~ sho + diasbp, data = whas)
whas9.6 <- coxph(Surv(los, dstat) ~ sho + bmi, data = whas)
whas9.7 <- coxph(Surv(los, dstat) ~ sho + cvd, data = whas)
whas9.8 <- coxph(Surv(los, dstat) ~ sho + afb, data = whas)
whas9.9 <- coxph(Surv(los, dstat) ~ sho + chf, data = whas)
whas9.10 <- coxph(Surv(los, dstat) ~ sho + av3, data = whas)
```

```
whas9.11 <- coxph(Surv(los, dstat) ~ sho + miord, data = whas)
whas9.12 <- coxph(Surv(los, dstat) ~ sho + mitype, data = whas)
whas9.13 <- coxph(Surv(los, dstat) ~ sho + year, data = whas)

BIC(whas9.1, whas9.2, whas9.3, whas9.4, whas9.5, whas9.6, whas9.7, whas9.8, whas9.9, whas9.10, whas9.11
```

```
##          df      BIC
## whas9.1   2 382.6642
## whas9.2   2 396.7218
## whas9.3   2 396.0587
## whas9.4   2 389.4640
## whas9.5   2 388.5351
## whas9.6   2 391.1468
## whas9.7   2 397.3442
## whas9.8   2 397.3014
## whas9.9   2 394.6917
## whas9.10  2 396.8511
## whas9.11  2 396.3923
## whas9.12  2 394.8969
## whas9.13  2 397.0672
```

Third Round of Fitting

```
whas9.1.1 <- coxph(Surv(los, dstat) ~ sho + age + gender, data = whas)
whas9.1.2 <- coxph(Surv(los, dstat) ~ sho + age + hr, data = whas)
whas9.1.3 <- coxph(Surv(los, dstat) ~ sho + age + sysbp, data = whas)
whas9.1.4 <- coxph(Surv(los, dstat) ~ sho + age + diasbp, data = whas)
whas9.1.5 <- coxph(Surv(los, dstat) ~ sho + age + bmi, data = whas)
whas9.1.6 <- coxph(Surv(los, dstat) ~ sho + age + cvd, data = whas)
whas9.1.7 <- coxph(Surv(los, dstat) ~ sho + age + afb, data = whas)
whas9.1.8 <- coxph(Surv(los, dstat) ~ sho + age + chf, data = whas)
whas9.1.9 <- coxph(Surv(los, dstat) ~ sho + age + av3, data = whas)
whas9.1.10 <- coxph(Surv(los, dstat) ~ sho + age + miord, data = whas)
whas9.1.11 <- coxph(Surv(los, dstat) ~ sho + age + mitype, data = whas)
whas9.1.12 <- coxph(Surv(los, dstat) ~ sho + age + year, data = whas)

BIC(whas9.1.1, whas9.1.2, whas9.1.3, whas9.1.4, whas9.1.5, whas9.1.6, whas9.1.7, whas9.1.8, whas9.1.9,
     whas9.1.10, whas9.1.11,  whas9.1.12)
```

```
##            df      BIC
## whas9.1.1   3 386.3049
## whas9.1.2   3 385.8816
## whas9.1.3   3 378.7440
## whas9.1.4   3 380.8626
## whas9.1.5   3 385.1454
## whas9.1.6   3 385.8329
## whas9.1.7   3 385.9228
## whas9.1.8   3 385.3418
## whas9.1.9   3 385.6011
## whas9.1.10  3 386.2457
```

```
## whas9.1.11  3 386.1579
## whas9.1.12  3 385.9805
```

Fourth Round of Fitting

```
whas9.1.3.1 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + gender, data = whas)
whas9.1.3.2 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + hr, data = whas)
whas9.1.3.3 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + diasbp, data = whas)
whas9.1.3.4 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + bmi, data = whas)
whas9.1.3.5 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + cvd, data = whas)
whas9.1.3.6 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + afb, data = whas)
whas9.1.3.7 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + chf, data = whas)
whas9.1.3.8 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + av3, data = whas)
whas9.1.3.9 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + miord, data = whas)
whas9.1.3.10 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + mitype, data = whas)
whas9.1.3.11 <- coxph(Surv(los, dstat) ~ sho + age + sysbp + year, data = whas)

BIC(whas9.1.3.1, whas9.1.3.2, whas9.1.3.3, whas9.1.3.4, whas9.1.3.5, whas9.1.3.6, whas9.1.3.7,
    whas9.1.3.8, whas9.1.3.9, whas9.1.3.10, whas9.1.3.11)
```

```
##              df      BIC
## whas9.1.3.1   4 382.2305
## whas9.1.3.2   4 380.9133
## whas9.1.3.3   4 382.0673
## whas9.1.3.4   4 381.6218
## whas9.1.3.5   4 382.2200
## whas9.1.3.6   4 382.1888
## whas9.1.3.7   4 381.2521
## whas9.1.3.8   4 381.6483
## whas9.1.3.9   4 382.4072
## whas9.1.3.10  4 381.9209
## whas9.1.3.11  4 381.2862
```

Citations

David W. Hosmer, Jr., et al. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data.* John Wiley & Sons, 2011.

"survminer: Survival Analysis and Visualization." *The Comprehensive R Archive Network*, cran.r-project.org/web/packages/survminer/readme/README.html.

"WHAS Survival." *John Wiley & Sons*, ftp://ftp.wiley.com/public/sci_tech_med/survival.