Predictions Survival Rates based on Vitals taken at Admission and Discharge/Death

January 20, 2020

Abstract

Shock is a serious life condition which can lead to death in a person. In Southern California, data was collected on patients twice when they were initially admitted and final vitals. Significant contributors to a patient's survival rate was their Shock Type, Mean Arterial Pressure, Mean Central Venous Pressure, Body Surface Index, Urinary Output, and Red Cell Index. Shock Type was a significant indicator on whether the patient will survive. Interaction terms were considered in the model selection, but ultimately not used due to stepAIC. Take a patient with Other Shock, MAP of 70, MCVP of 100, BSI of 150, UO of 75, RCI of 150. This patient has 17676.65 odds more likely to die.

1 Introduction

Death by shock is a life threatening condition where the body does not get enough blood flow. There was data collected in Southern California on patients when they were being admitted and final measurements. There were 43 patients who succumbed to their conditions and 69 who survived. Specifically, 40 of the deaths were due to some category of shock. Only 3 deaths were not of shock. This analysis will determine what the best indicators are to predict the patient's survival rate based on measurements taken when the patient got admitted and final vitals measured. Logistic Regression will be applied because the response is binary. There is a hypothesis that shock type will greatly affect whether a patient survives.

Table 1: Survival by Sex Table

	Died	Survived
Female	26	28
Male	17	41

2 Methods

There were 112 critically ill patients that were admitted to an Emergency room. Vitals were taken upon admission and taken again right before death or discharge. Record showed whether the vitals were taken upon admission or as final measurements. ID was the patient's identification. Age was the age of the patient in years. Sex of the patient was either male (1) or female (2). Height is the height of the patient in cm's. Survive was whether the patient had survived (0) or died (1). Shock type was the patient's status and what kind of shock they were in: Non-Shock meant the patient did not have any shock. Hypovolemic Shock is losing severe amount of blood. Cardiogenic Shock is when a patient's heart cannot pump enough blood to meet the body's demand. Bacterial Shock is a complication of certain bacterial infections. Neurogenic Shock is irregular blood circulation usually leading to low blood pressure often caused by trauma or injury to the

spinal cord. SBP was their systolic blood pressure measured in mmHg. MAP was the patient's mean arterial pressure measured in mmHg. HR was the patient's heart rate in beats/mins. DBP is diastolic blood pressure measured in mmHg. MCVP is mean central venous water pressure measured in cm. BSI is the patient's body surface index measured in m^2. CI is cardiac index measured in liters/(min * m^2). AT is appearance time of physical symptoms in seconds. MCT is mean circulation time measured in seconds. UO is urinary output is measured mL/hr. PVI is Plasma volume index is measured in mL/kg. RCI is the red cell index measured in mL/kg. HG is hemoglobin is gm/100mL. HCT is hematocrit is measured in percent. Record is whether the initial or final measurement was taken. The analysis will be conducted using R/RStudio.

Table 2: Survival by Shock Type Table

	Bacterial	Cardiogenic	Hypovolemic	Neurogenic	Non-Shock	Other
Survived	9	10	7	9	31	3
Died	6	10	10	7	3	7

3 Results

3.1 Exploratory Data Analysis

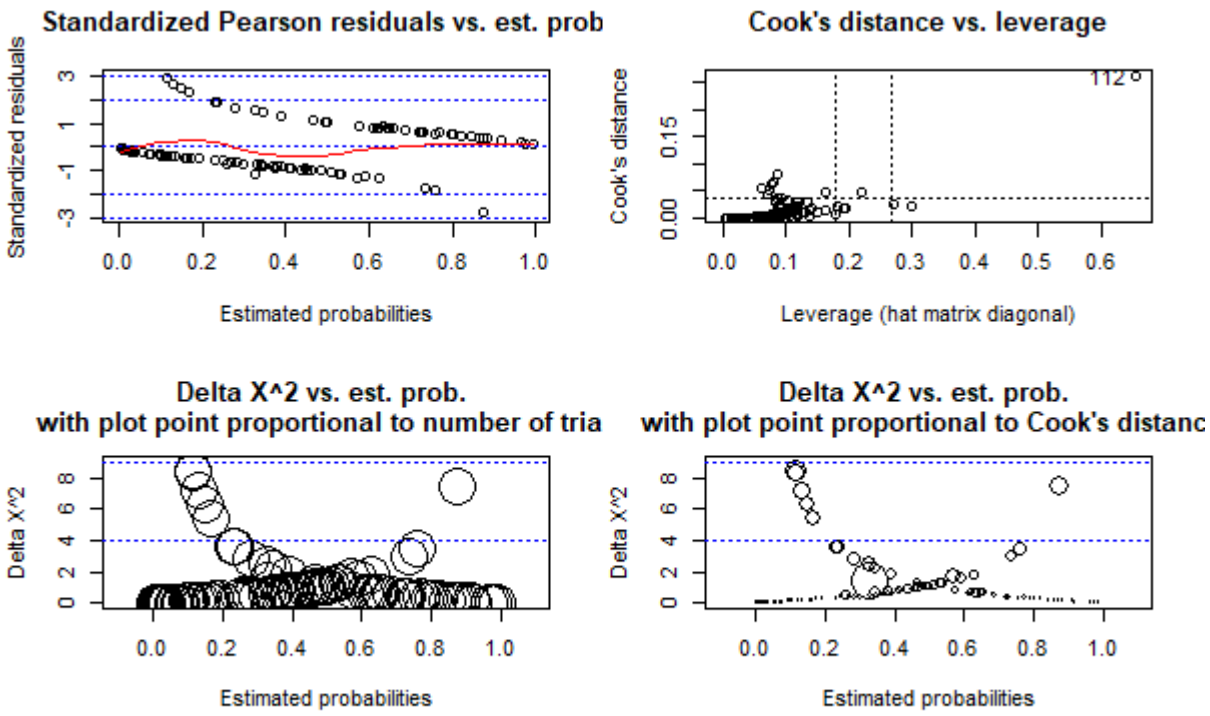
In table 1, there was a surprising discovery that men survived at a rate more than double compared to women. In table 2 showed that the largest group that survived was those who were not in shock. In these tables showed that males with a non-shock condition will likely survive. By using the function stepAIC on all possible models, the model with the least of amount of penalties still had non significant terms. Manually, the terms that were not significant by p-values were dropped one-by-one out of the model, but dropped too many covariates including shock type. Eventually, the model contained six predictor variables that were not highly correlated with each other shown in figure 2 in the appendix. The correlation plot of all variables showed that the variables SBP, DBP, and MAP were highly correlated. MCT was highly correlated to AT. HCT was highly correlated to HG. MCT and AT were inversely correlated to CI. With dimension reduction and model selection, the model was lowered to 6 covariates. All variables except for survive, sex, shock type, and record were continuous.

3.2 Model Fitting/Inferences

Interpreting an odds ratio is is not too difficult. Any odds ratio at 1 does not in influence the response. Odds ratios that are between 0 and 1 are (1 - odds ratio) * 100% less odds of dying. But with an odds ratio greater than 1 and less than 2, then the patient has (odds ratio - 1) * 100% more odds of dying. Patients that had Cardiogenic Shock had 2.53 more likely odds in dying. Patients with Hypovolemic Shock had 3.52 more likely odds in dying. Patients with Neurogenic Shock had 2.28 more likely odds in dying. Patients with no shock had 64% decrease in odds of dying. Patients with other shock types had 12.21 more likely odds of dying. Patients with every 1 unit MAP increase, decreases their odds of dying by 3%. Patients with every 1 unit MCVP increase, increases their odds of dying by 1%. Patients with every one unit increase of BSI, their odds of dying decrease by 3%. Patients with every 1 unit of UO increases decreases their odds of dying by 1%. Every increase of RCI in a patient, decreases their odds of dying by 1%. With this model, a patient's odds of death can be determined. Based on table 3, take a patient with Other Shock, MAP of 70, MCVP of 100, BSI of 150, UO of 75, RCI of 150. This patient has 17676.65 odds more likely to die.

In the diagnostic plots figure 1 showed an influential point of patient 112. Looking at the Delta X^2 plot showed no significant impact to the model since all probabilities were contained under the 2^{nd} dotted line. Applying the Homser-Lemeshow test with 3 bins, the p-value is 0.13 means fail to reject the chosen model. The other thing of note is the pattern in the plot of the Standardized Pearson residuals vs estimated probabilities which seems to have 2 parallel lines. This is due to the limited number of observations and the

Figure 1: Diagnostic Plot of the Final Model



Deviance/df = 0.98; GOF thresholds: 2 SD = 1.28, 3 SD = 1.42

data set itself. The confusion matrix table 4 had a sensitivity of 71% and a specificity of 74%. It would be ideal to have both metrics above 90% or at least optimally above 80%.

Table 3: Odds Ratios with 95% CI Odds Ratios

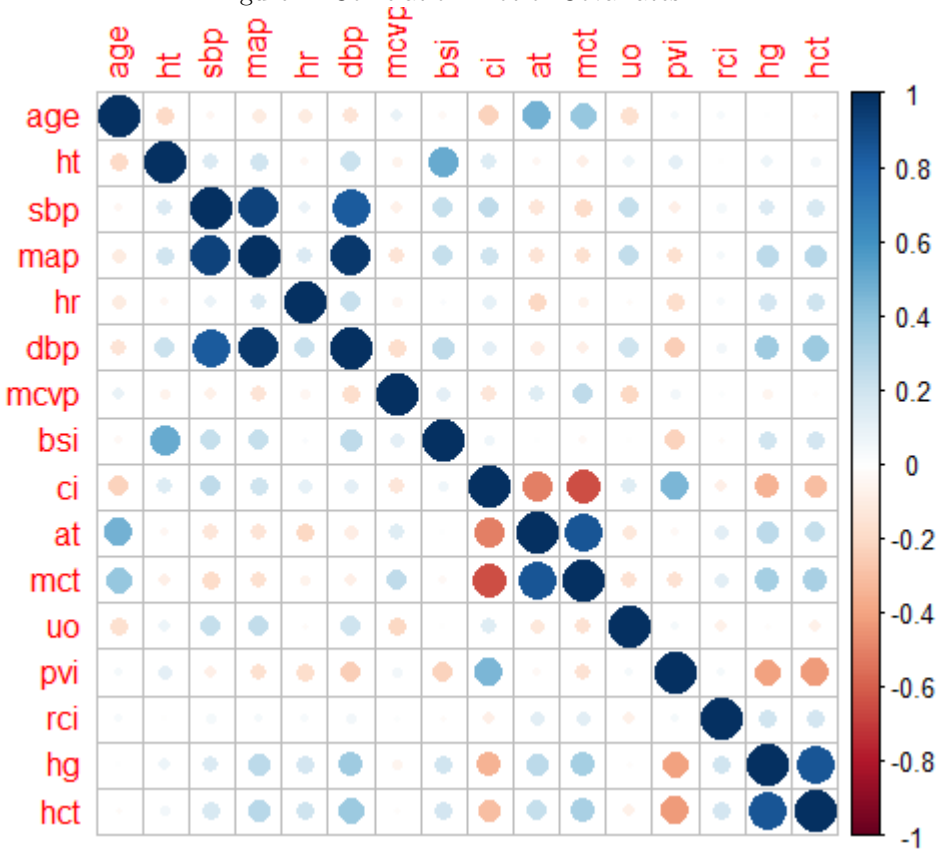
	Coefficients	Odds Ratio	2.5 %	97.5 %
(Intercept)	6.14	462.70	3.75	99681.27
Cardiogenic Shock	0.93	2.54	0.52	13.59
Hypovolemic Shock	1.26	3.52	0.63	22.36
Neurogenic Shock	0.82	2.28	0.42	13.44
Non-Shock	-1.00	0.36	0.05	2.16
Other Shock	2.58	13.21	1.56	160.29
Mean Arterial Pressure	-0.03	0.97	0.94	0.99
Mean Central Venous Pressure	0.01	1.01	1.00	1.03
Body Surface Index	-0.03	0.97	0.94	1.00
Urinary Output	-0.006	0.99	0.98	1.00
Red Cell Index	-0.005	0.99	0.99	1.00

4 Conclusion

This analysis has shown that Shock Type, Mean Arterial Pressure, Mean Central Venous Pressure, Body Surface Index, Urinary Output, and Red Cell Index are the best predictors in determining whether a patient is going to die. The Shock Type is a great indicator on whether the patient survives or not. As an example, take a patient with Other Shock, MAP of 70, MCVP of 100, BSI of 150, UO of 75, RCI of 150. This patient has 17676.65 odds more likely to die. When the patient is not in a shocked disposition, then rate of survival dramatically increases. There were variables which helped the survival of the patient. Covariates such as Mean arterial Pressure, Mean Central Venous Pressure, Body Surface Index, Urinary Output, Red Cell Index helped with survival rates. The greater in magnitude in those variables resulted in better odds of the patients surviving.

This data set had several limitations. One limitation was the small amount of observations and paired measurements per observation. Interaction terms were considered, but not used because models would have non-converging algorithms full of NA's. There was no time variable where survival or time series analysis could have supplemented the analysis. Further analysis would include multiple regression on a subset which parse the data by record and sex to see if any significant information could be teased.

Figure 2: Correlation Plot of Covariates



Appendix A: Auxiliary Graphics and Tables

Table 4: Confusion Matrix

	actual survive	actual death
predicted survive	49	11
predicted death	20	32

Appendix B: R Code

```r
library(readxl)  # To read xlsx
library(readr)   # Fast csv write
library(MASS) #stepAIC
library(car) #vif
library(xtable)

#converting xlsx to csv
setwd("C:/Users/Kelso Quan/Documents/DAR")
df <- read_excel("DATA-FILEsp2020.xlsx")
write_csv(df, path="patient.csv")
data <- read.csv("patient.csv", header = F)
sum(is.na(data))
###tidy###
colnames(data) <-  c("id","age", "ht", "sex", "survive", "shock_type", "sbp",
                     "map", "hr", "dbp", "mcvp", "bsi", "ci", "at", "mct", "uo",
                     "pvi", "rci", "hg", "hct", "record")
data[data$survive == 1, "survive"] <- 0 #0 survived
data[data$survive == 3, "survive"] <- 1 #1 died
as.factor(data$survive)
data$shock_type <- ifelse(data$shock_type==2,"Non-Shock",
                          ifelse(data$shock_type==3,"HypoVolemic",
                                 ifelse(data$shock_type==4,"Cardiogenic",
                                        ifelse(data$shock_type==5,"Bacterial",
                                               ifelse(data$shock_type==6,"Neurogenic",
                                                      "Other")))))
as.factor(data$sex)
attach(data)
pre <- subset(data ,record == 1)

#eda#

hist(age)
hist(ht)
hist(sbp)
hist(pre$sbp)
hist(map)
hist(pre$map)
hist(hr)
hist(pre$hr)
hist(dbp)
hist(pre$dbp)
hist(mcvp)
hist(pre$mcvp)
hist(log(mcvp))
hist(bsi)
hist(pre$bsi)
hist(ci)
hist(pre$ci)
hist(log(ci))
hist(at)
hist(pre$at)
hist(log(at)) #normalize
hist(mct)
hist(pre$mct)
hist(log(mct))
hist(uo)
hist(pre$uo)
hist(log(uo))
hist(pvi)
hist(pre$pvi)
hist(log(pvi))#normalize
hist(log(pre$pvi))
hist(rci)
```

```r
hist(pre$rci)
hist(log(rci))
hist(hg)
hist(pre$hg)
hist(hct)
hist(pre$hct)


model <- glm(survive ~ age + ht + sex + shock_type + sbp + map+
               hr + dbp + mcvp+ bsi+ ci+ at+ mct + uo + pvi +
               rci + hg + hct , family = binomial(link = "logit"), data = pre)
summary(model)
stepAIC(model)
final <- glm(formula = survive ~ shock_type + map + mcvp + bsi + uo +
               rci, family = binomial(link = "logit"), data = pre)
summary(final)
coef(final)
exp(coef(final))
exp(final$coefficients)
exp(confint(final))
xtable(exp(confint(final)))


threshold <- 43/112
predict.final <- ifelse(predict(final,type = "response")>threshold,1,0)
actual.final <- final$y
confused <- table(predict.final,actual.final)
confused
xtable(confused)

cont <- data.frame(age, ht, sbp,map,hr,dbp,mcvp,bsi,ci,at,mct,uo,pvi,rci,hg,hct)
corrplot(cor(cont))
xtable(corrplot(cor(cont)))

final.cont <- data.frame(map,bsi,uo,mcvp,rci)
corrplot(cor(final.cont))
####
one.fourth.root=function(x){
  x^0.25
}
source("examine.logistic.reg.R")
# Consider model
dat.glm <- glm(formula = survive ~ shock_type + map + mcvp + bsi + uo +
                 rci, family = binomial, data = pre)
dat.mf <- model.frame(dat.glm)
## Covariate pattern: too many EVPs!
w <- aggregate(formula = survive ~ shock_type + map + mcvp + bsi + uo +
                 rci, data = pre, FUN = sum)
n <- aggregate(formula = survive ~ shock_type + map + mcvp + bsi + uo +
                 rci, data = pre, FUN = length)
w.n <- data.frame(w, trials = n$survive, prop = round(w$survive/n$survive,2))
dim(w.n)
#[1] 112 9

# Create EVPs by binning continuous covariates
g = 8 # number of categories
mcvp_interval = cut(mcvp, quantile(mcvp, 0:g/g), include.lowest = TRUE)  # Creates factor
    with levels 1,2,...,g
levels(mcvp_interval)

# Diagnostic plots
w <- aggregate(formula = survive ~ shock_type + map + mcvp + bsi + uo +
                 rci, data = pre, FUN = sum)
n <- aggregate(formula = survive ~ shock_type + map + mcvp + bsi + uo +
                 rci, data = pre, FUN = length)
```

```
w.n <- data.frame(w, trials = n$survive, prop = round(w$survive/n$survive,2))
mod.prelim1 <- glm(formula = survive/trials ~ mcvp+rci+map+shock_type+bsi,
                   family = binomial(link = logit), data = w.n, weights = trials)
save1 = examine.logistic.reg(mod.prelim1, identify.points=T, scale.n=one.fourth.root, scale
    .cookd=sqrt)

# Evaluation of EVPs for potential outlying sets of points
w.n.diag1 = data.frame(w.n, pi.hat=round(save1$pi.hat, 2), std.res=round(save1$stand.resid,
     2),
                       cookd=round(save1$cookd, 2), h=round(save1$h, 2))
p = length(mod.prelim1$coef) # number of parameters in model (# coefficients)
ck.out = abs(w.n.diag1$std.res)>2 | w.n.diag1$cookd>4/nrow(w.n) | w.n.diag1$h > 3*p/nrow(w.
    n)
extract.EVPs = w.n.diag1[ck.out, ]
extract.EVPs

#HL test
source("HLTest.R")
HLTest(mod.prelim1, 3)
```

Listing 1: Survival Rates