What factors determine survival among Cardiovascular Heart Disease (CHD) Patients?

Yordanos Alemu, Rabab Mohammed, Kovenda Mubale

1/16/2021

## Abstract:

This data set from Plos One, is about Cardiovascular Heart Disease (CHD). The study is based on 299 patients of Cardiovascular Heart Disease (CHD) comprising of 105 women and 194 men. All the patients were more than 40 years old, having left ventricular systolic dysfunction and falling in NYHA class III and IV. Cardiovascular Heart Disease (CHD) was diagnosed by cardiac echo report or notes written by physician. The information related to risk factors were taken from blood reports while smoking status and blood pressure were taken from physician's notes. Cardiovascular Heart Disease (CHD) is now top reason causing 31% of deaths globally. Pakistan is also included in the list of countries where prevalence of CHD is increasing significantly. According to report by Al-Shifa hospital, 33% of Pakistani population above 45 has hypertension, 25% of patients over 45 years suffer diabetes mellitus, and CHD deaths in Pakistan has reached about 200,000 per year i.e. 410/100,000 of the population). All this results in increased prevalence of heart failure. Rate of heart failure patients in Pakistan is estimated to be 110 per million. Rising stress of economic and social issues in the modern era, greasy food with little exercise results towards increased prevalence of heart failure in Pakistan. The main objective of this study is to determine the factors that are critical for survival among Cardiovascular Heart Disease (CHD) Patients by using data from Faisalabad (third most populous city of Pakistan).

library(readr)  
survival <- read\_csv("S1Data\_fixed.csv")

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## TIME = col\_double(),  
## Survival = col\_double(),  
## Gender = col\_double(),  
## Smoking = col\_double(),  
## Diabetes = col\_double(),  
## BP = col\_double(),  
## Anaemia = col\_double(),  
## Age = col\_double(),  
## EF = col\_double(),  
## Sodium = col\_double(),  
## `Serum Creatinine` = col\_double(),  
## platelets = col\_double(),  
## CPK = col\_double()  
## )

attach (survival)  
head(survival)

## # A tibble: 6 x 13  
## TIME Survival Gender Smoking Diabetes BP Anaemia Age EF Sodium  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 97 1 0 0 0 0 1 43 2 1  
## 2 180 1 1 1 1 0 1 73 0 3  
## 3 31 0 1 1 0 1 0 70 0 0  
## 4 87 1 1 0 0 0 1 65 0 3  
## 5 113 1 1 0 0 0 0 64 2 1  
## 6 10 0 1 0 0 0 1 75 0 1  
## # … with 3 more variables: `Serum Creatinine` <dbl>, platelets <dbl>, CPK <dbl>

na.omit(survival)

## # A tibble: 299 x 13  
## TIME Survival Gender Smoking Diabetes BP Anaemia Age EF Sodium  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 97 1 0 0 0 0 1 43 2 1  
## 2 180 1 1 1 1 0 1 73 0 3  
## 3 31 0 1 1 0 1 0 70 0 0  
## 4 87 1 1 0 0 0 1 65 0 3  
## 5 113 1 1 0 0 0 0 64 2 1  
## 6 10 0 1 0 0 0 1 75 0 1  
## 7 250 1 1 1 0 0 0 70 1 1  
## 8 27 0 1 0 1 1 0 94 1 0  
## 9 87 1 1 0 0 1 0 75 1 1  
## 10 87 1 1 1 0 0 0 80 0 3  
## # … with 289 more rows, and 3 more variables: `Serum Creatinine` <dbl>,  
## # platelets <dbl>, CPK <dbl>

survival$index <-1:nrow(survival)  
head(survival)

## # A tibble: 6 x 14  
## TIME Survival Gender Smoking Diabetes BP Anaemia Age EF Sodium  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 97 1 0 0 0 0 1 43 2 1  
## 2 180 1 1 1 1 0 1 73 0 3  
## 3 31 0 1 1 0 1 0 70 0 0  
## 4 87 1 1 0 0 0 1 65 0 3  
## 5 113 1 1 0 0 0 0 64 2 1  
## 6 10 0 1 0 0 0 1 75 0 1  
## # … with 4 more variables: `Serum Creatinine` <dbl>, platelets <dbl>,  
## # CPK <dbl>, index <int>

## Data Characteristic:

Response variable:

Survival - Cardiovascular Heart Disease (CHD) Patients that were still alive after the follow up period by the physicians. If the patient was alive by the follow up date: 1 and if not: 0.

Predictor variables:

Age, serum sodium, serum creatinine, gender, smoking, Blood Pressure (BP), Ejection Fraction (EF), anemia, platelets, Creatinine Phosphokinase (CPK) and diabetes. These are considered as potential variables explaining mortality caused by Cardiovascular Heart Disease (CHD).

Age and Time are continuous variables. EF, serum sodium, platelets, CPK and serum creatinine were taken as categorical variables.

Time is follow up time was 4 to 285 days with an average of 130 days.

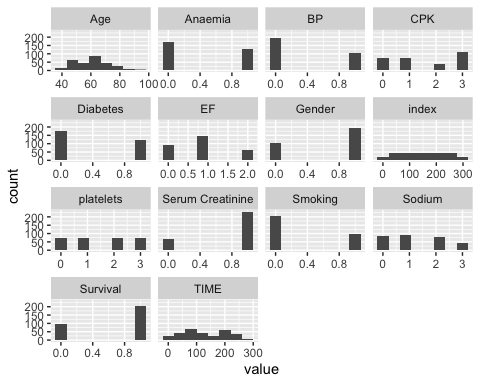
EF was divided into three levels (i.e. EF<30 (coded as 0), 3045 (coded as 2)) and platelets was also divided into three level on the basis of quartiles.

Sodium was also coded into three levels based on the quartiles

Serum Creatinine greater than its normal level (1.5) is an indicator of renal dysfunction. Its effect on mortality was studied as creatinine >1.5 (coded as 0) vs <1.5 (coded as 1).

Anemia in patients was assessed by their haematocrit level. The patients with haematocrit less than 36 (minimum normal level of haematocrit) were taken as anemic.

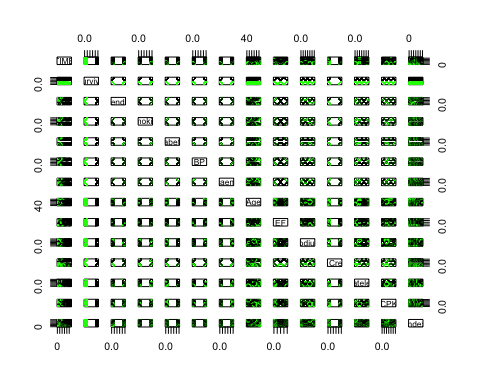
library (ggplot2)  
library (tidyr)  
  
ggplot(gather(survival), aes(value)) +   
 geom\_histogram(bins = 8) +   
 facet\_wrap(~key, scales = 'free\_x')



All the predictor variables are not extremely right or left skewed. We think that all the variables are categorical variables expect for Age and Time.

# Scatterplot matrix of columns:

# Scatterplot matrix of columns   
pairs (survival, col=ifelse (Survival==0, "green", "black") )



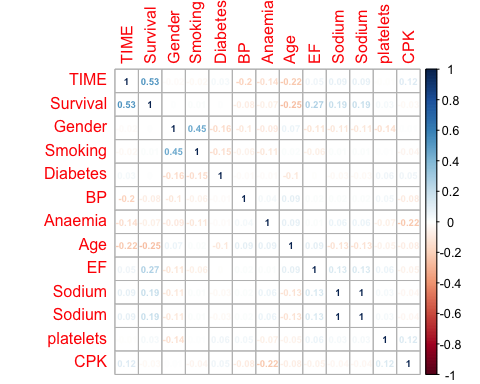
#pairs (survival[,2:10], col= EF)

The Explanatory Analysis shows that all our variables are categorical except for Time and Age.

transcorr = cor(survival [, c(1:10, 10, 12:13)], use= "complete.obs")   
library("corrplot")

## corrplot 0.84 loaded

## corrplot 0.84 loaded  
corrplot (transcorr, method = "number", number.cex=0.6)



The correlation matrix shows that the variable Time and survival are highly correlated (0.53). Gender and Smoking have a high correlation as well(0.45). Survival and Serum Creatinine appears to have a moderate correlation(0.37).

library(MASS)   
table1= xtabs(~Survival+EF)  
table1

## EF  
## Survival 0 1 2  
## 0 51 31 14  
## 1 42 115 46

93 patients had an Ejection Fraction of less than 30 %, while 146 have an Ejection Fraction of 30 to 45 % and only 60 have an Ejection Fraction higher than 45% . Of all the patients that had an ejection fraction less than 30 %, 55 % died of cardio-vascular heart disease. That is 32% more than the patients who died that had an ejection fraction higher than 45%. Patients with an ejection fraction between 30 and 45 % had the highest survival rate at 79 %. You have a higher chance of dying if your ejection fraction is less than 30 %.

table2= xtabs(~Survival+Gender)  
table2

## Gender  
## Survival 0 1  
## 0 34 62  
## 1 71 132

Odds of surviving . Male 132/194 = 0.680 . Female 71/105 = 0.676 .... Odds Ratio for Male VS Female -> 0.680/0.676 = 1.0059

Female patients are 1.0059 times more likely to survive cardio-vascular heart diseases than male patients. This essentially means the chances of survival for male and female patients is approximately 1 to 1.

tabel4=xtabs(~Survival+Diabetes)  
tabel4

## Diabetes  
## Survival 0 1  
## 0 56 40  
## 1 118 85

Odds of not surviving . Diabetic 40/125 = 0.32 . Non-diabetic 56/174 = 0.32 .... Odds Ratio for Diabetic VS Non-Diabetic -> 0.32/0.32 = 1

This essentially means the chances of not surviving for Diabetic and non-Diabetic patients is 1 to 1.

tabel7=xtabs(~Survival+Sodium)  
tabel7

## Sodium  
## Survival 0 1 2 3  
## 0 42 24 19 11  
## 1 41 70 61 31

Urinary Sodium to creatine ratio or Sodium levels 0 to 3 represent the quartiles of the collected Sodium ratios, 0 being less 25th quartile and 3 being greater than 75th quartile. We can see that when the patient's Sodium to Creatine ratio is less than the 25th quartile they have less than 50% chance of surviving cardio-vascular heart disease. While patients with Sodium to Creatine ratio greater than the 25th quartile all have higher than 70% chance of surviving cardio-vascular heart disease.

# Initial Model:

order1\_fit1 <- glm (Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME + Gender + Smoking + BP + Anaemia + Age + Sodium + platelets + CPK, family = binomial)   
summary(order1\_fit1)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + `Serum Creatinine` +   
## Diabetes + TIME + Gender + Smoking + BP + Anaemia + Age +   
## Sodium + platelets + CPK, family = binomial)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4184 -0.4308 0.1899 0.5562 2.7121   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.358485 1.308658 -1.038 0.299235   
## as.factor(EF)1 1.836955 0.422611 4.347 1.38e-05 \*\*\*  
## as.factor(EF)2 1.857695 0.518107 3.586 0.000336 \*\*\*  
## `Serum Creatinine` 1.678121 0.429314 3.909 9.27e-05 \*\*\*  
## Diabetes -0.192139 0.359440 -0.535 0.592961   
## TIME 0.022433 0.003194 7.024 2.15e-12 \*\*\*  
## Gender 0.553463 0.436964 1.267 0.205295   
## Smoking -0.135759 0.430672 -0.315 0.752590   
## BP -0.009470 0.368440 -0.026 0.979495   
## Anaemia -0.049583 0.379851 -0.131 0.896146   
## Age -0.046199 0.016617 -2.780 0.005432 \*\*   
## Sodium 0.292238 0.184413 1.585 0.113036   
## platelets 0.114345 0.165775 0.690 0.490344   
## CPK -0.397067 0.161646 -2.456 0.014034 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 209.93 on 285 degrees of freedom  
## AIC: 237.93  
##   
## Number of Fisher Scoring iterations: 6

exp (order1\_fit1$coefficients)

## (Intercept) as.factor(EF)1 as.factor(EF)2 `Serum Creatinine`   
## 0.2570498 6.2773944 6.4089462 5.3554847   
## Diabetes TIME Gender Smoking   
## 0.8251924 1.0226867 1.7392656 0.8730532   
## BP Anaemia Age Sodium   
## 0.9905751 0.9516266 0.9548522 1.3394224   
## platelets CPK   
## 1.1211391 0.6722891

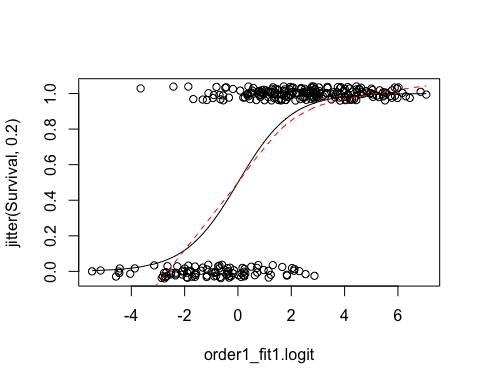
exp (confint (order1\_fit1))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.01872896 3.2432977  
## as.factor(EF)1 2.80878250 14.8482312  
## as.factor(EF)2 2.40283407 18.5011624  
## `Serum Creatinine` 2.35640903 12.7988731  
## Diabetes 0.40569020 1.6706351  
## TIME 1.01669786 1.0295687  
## Gender 0.74725273 4.1808057  
## Smoking 0.37153732 2.0267428  
## BP 0.48198078 2.0566115  
## Anaemia 0.45160198 2.0164303  
## Age 0.92319656 0.9856701  
## Sodium 0.93674051 1.9377001  
## platelets 0.81029581 1.5569596  
## CPK 0.48416180 0.9155519

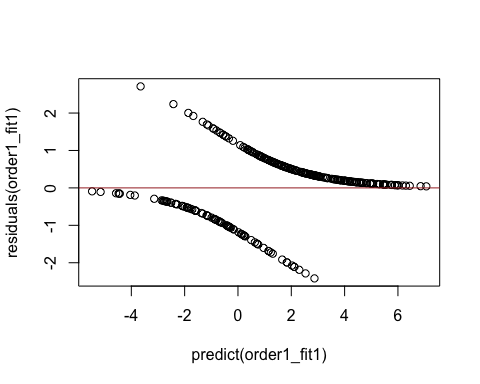
The model shows that patients having an ejection fraction between 30 to 45 % and higher 45 % are significantly different from patients who have an ejection fraction of less than 30 %. The model also shows that Serum Creatinine, Time, Age and CPK are significant predictors for predicting the survival of cardio-vascular heart disease. For every unit increase in Serum Creatinine the odds of surviving cardio-vascular disease go up by 5.3554847-folds. The odds of surviving cardio-vascular heart disease for male patients vs female patients is not only 1 to 1 as we predicted earlier but essentially ranges from 0.74725273 to 4.1808057 -folds.

# Make a plot similar to Y vs Y-hat for linear regression  
order1\_fit1.logit = predict (order1\_fit1)  
plot (jitter (Survival, 0.2) ~ order1\_fit1.logit, data=survival)  
# Add the overall fitted logistic curve to the plot, and a lowess fit  
pihat.order1\_fit1 = predict (order1\_fit1, type='response')  
pihat.ord = order (pihat.order1\_fit1)  
lines (order1\_fit1.logit [pihat.ord], pihat.order1\_fit1 [pihat.ord])  
lines (lowess (order1\_fit1.logit [pihat.ord], pihat.order1\_fit1 [pihat.ord]), col='red', lty=2)



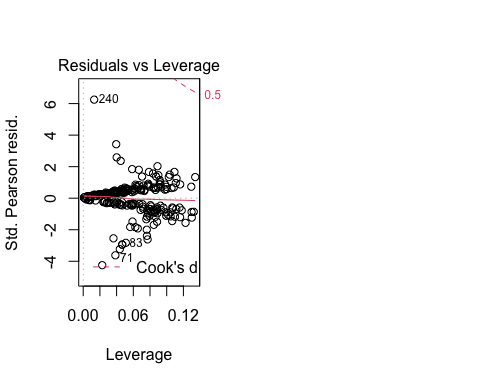
The jittered response vs. predicted values with the fitted logistic curve and a lowess fit shows that the model fits the data approximately well since the solid and the jittered line are approximately similar in shape.

plot(predict(order1\_fit1),residuals(order1\_fit1))  
abline(h=0,lty=1,col="brown")



The linearity condition is satisfied because the plot is flat and right on zero.

par (mfrow=c(1,2))   
plot (order1\_fit1, which=c(5))



There are no points with high leverage or high Cook's Distance therefore those influence measures do not appear on the plot. The point index 240 has a standardized residual around 6. We will see whether it is an influential point .

# Inference Analysis to confirm that there is no outliers:

survival$Residual = round (order1\_fit1$residuals, 4)  
survival$leverage = round (hatvalues(order1\_fit1), 4)  
survival$rstudent = round (rstudent(order1\_fit1), 4)

Pull out the cases with high leverage and large residuals.

high.levg.resd = survival [survival$leverage > 3\*(order1\_fit1$rank) / (order1\_fit1$rank + order1\_fit1$df.residual) |abs (order1\_fit1$residuals) > 30 , ]  
  
high.levg.resd[order(-high.levg.resd$rstudent),][c(1,9:17)]

## # A tibble: 1 x 10  
## TIME EF Sodium `Serum Creatini… platelets CPK index Residual leverage  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <int> <dbl> <dbl>  
## 1 30 0 1 0 3 3 240 39.6 0.0131  
## # … with 1 more variable: rstudent <dbl>

Although index 240 has a large residual value, however, its 0.0131 leverage value is not bigger than the Leverage Cutoff value at 0.13043478 (3(k+1)/n). Therefore, this point is an influential point.

We will see whether there are any other influential points that do not have high residuals as index 240.

high.Leverage = survival [survival$leverage > 0.13043478 & is.na(survival$leverage) == FALSE, ]  
high.Leverage[order(-high.Leverage$leverage),][c(1,9:17)]

## # A tibble: 2 x 10  
## TIME EF Sodium `Serum Creatini… platelets CPK index Residual leverage  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <int> <dbl> <dbl>  
## 1 94 2 2 0 3 0 242 2.55 0.134  
## 2 172 0 3 0 3 3 222 -1.65 0.132  
## # … with 1 more variable: rstudent <dbl>

There are only 2 points with leverage values higher than the Leverage Cutoff value at 0.13043478 (3(k+1)/n). The point index 242 has the highest leverage value of 0.134.

car::vif(order1\_fit1)

## GVIF Df GVIF^(1/(2\*Df))  
## as.factor(EF) 1.328792 2 1.073654  
## `Serum Creatinine` 1.166354 1 1.079979  
## Diabetes 1.044061 1 1.021793  
## TIME 1.280348 1 1.131525  
## Gender 1.464522 1 1.210174  
## Smoking 1.343264 1 1.158993  
## BP 1.063805 1 1.031409  
## Anaemia 1.172183 1 1.082674  
## Age 1.208652 1 1.099387  
## Sodium 1.122267 1 1.059371  
## platelets 1.088017 1 1.043081  
## CPK 1.254166 1 1.119896

All of the VIF are less than 5. Hence, multi-colinearity is low in our model. We will be doing stepwise regression to find the best model.

# Model Selection:

step.AIC = step (order1\_fit1, direction = "both")

## Start: AIC=237.93  
## Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME +   
## Gender + Smoking + BP + Anaemia + Age + Sodium + platelets +   
## CPK  
##   
## Df Deviance AIC  
## - BP 1 209.93 235.93  
## - Anaemia 1 209.94 235.94  
## - Smoking 1 210.03 236.03  
## - Diabetes 1 210.21 236.21  
## - platelets 1 210.40 236.40  
## - Gender 1 211.56 237.56  
## <none> 209.93 237.93  
## - Sodium 1 212.49 238.49  
## - CPK 1 216.35 242.35  
## - Age 1 218.19 244.19  
## - `Serum Creatinine` 1 226.47 252.47  
## - as.factor(EF) 2 234.52 258.52  
## - TIME 1 287.94 313.94  
##   
## Step: AIC=235.93  
## Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME +   
## Gender + Smoking + Anaemia + Age + Sodium + platelets + CPK  
##   
## Df Deviance AIC  
## - Anaemia 1 209.94 233.94  
## - Smoking 1 210.03 234.03  
## - Diabetes 1 210.22 234.22  
## - platelets 1 210.40 234.40  
## - Gender 1 211.59 235.59  
## <none> 209.93 235.93  
## - Sodium 1 212.49 236.49  
## + BP 1 209.93 237.93  
## - CPK 1 216.39 240.39  
## - Age 1 218.19 242.19  
## - `Serum Creatinine` 1 226.64 250.64  
## - as.factor(EF) 2 234.56 256.56  
## - TIME 1 289.64 313.64  
##   
## Step: AIC=233.95  
## Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME +   
## Gender + Smoking + Age + Sodium + platelets + CPK  
##   
## Df Deviance AIC  
## - Smoking 1 210.04 232.04  
## - Diabetes 1 210.24 232.24  
## - platelets 1 210.45 232.45  
## - Gender 1 211.63 233.63  
## <none> 209.94 233.94  
## - Sodium 1 212.49 234.49  
## + Anaemia 1 209.93 235.93  
## + BP 1 209.94 235.94  
## - CPK 1 216.50 238.50  
## - Age 1 218.32 240.32  
## - `Serum Creatinine` 1 226.69 248.69  
## - as.factor(EF) 2 234.69 254.69  
## - TIME 1 291.15 313.15  
##   
## Step: AIC=232.04  
## Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME +   
## Gender + Age + Sodium + platelets + CPK  
##   
## Df Deviance AIC  
## - Diabetes 1 210.32 230.32  
## - platelets 1 210.51 230.51  
## - Gender 1 211.73 231.73  
## <none> 210.04 232.04  
## - Sodium 1 212.54 232.54  
## + Smoking 1 209.94 233.94  
## + Anaemia 1 210.03 234.03  
## + BP 1 210.03 234.03  
## - CPK 1 216.52 236.52  
## - Age 1 218.49 238.49  
## - `Serum Creatinine` 1 226.72 246.72  
## - as.factor(EF) 2 234.93 252.93  
## - TIME 1 291.30 311.30  
##   
## Step: AIC=230.32  
## Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Gender +   
## Age + Sodium + platelets + CPK  
##   
## Df Deviance AIC  
## - platelets 1 210.75 228.75  
## - Gender 1 212.14 230.14  
## <none> 210.32 230.32  
## - Sodium 1 212.96 230.96  
## + Diabetes 1 210.04 232.04  
## + Smoking 1 210.24 232.24  
## + Anaemia 1 210.31 232.31  
## + BP 1 210.32 232.32  
## - CPK 1 216.84 234.84  
## - Age 1 218.53 236.53  
## - `Serum Creatinine` 1 227.02 245.02  
## - as.factor(EF) 2 234.97 250.97  
## - TIME 1 291.52 309.52  
##   
## Step: AIC=228.75  
## Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Gender +   
## Age + Sodium + CPK  
##   
## Df Deviance AIC  
## - Gender 1 212.34 228.34  
## <none> 210.75 228.75  
## - Sodium 1 213.29 229.29  
## + platelets 1 210.32 230.32  
## + Diabetes 1 210.51 230.51  
## + Smoking 1 210.69 230.69  
## + Anaemia 1 210.71 230.71  
## + BP 1 210.75 230.75  
## - CPK 1 216.93 232.93  
## - Age 1 219.29 235.29  
## - `Serum Creatinine` 1 227.70 243.70  
## - as.factor(EF) 2 236.02 250.02  
## - TIME 1 291.55 307.55  
##   
## Step: AIC=228.34  
## Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Age +   
## Sodium + CPK  
##   
## Df Deviance AIC  
## <none> 212.34 228.34  
## - Sodium 1 214.42 228.42  
## + Gender 1 210.75 228.75  
## + Diabetes 1 211.98 229.98  
## + platelets 1 212.14 230.14  
## + Anaemia 1 212.23 230.23  
## + Smoking 1 212.23 230.23  
## + BP 1 212.30 230.30  
## - CPK 1 217.80 231.80  
## - Age 1 220.17 234.17  
## - `Serum Creatinine` 1 230.15 244.15  
## - as.factor(EF) 2 236.94 248.94  
## - TIME 1 292.78 306.78

summary(step.AIC)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + `Serum Creatinine` +   
## TIME + Age + Sodium + CPK, family = binomial)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4321 -0.4395 0.1993 0.5316 2.7943   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.154578 1.171306 -0.986 0.324271   
## as.factor(EF)1 1.831546 0.418696 4.374 1.22e-05 \*\*\*  
## as.factor(EF)2 1.770562 0.502666 3.522 0.000428 \*\*\*  
## `Serum Creatinine` 1.694108 0.417850 4.054 5.03e-05 \*\*\*  
## TIME 0.022141 0.003103 7.136 9.59e-13 \*\*\*  
## Age -0.043401 0.015976 -2.717 0.006596 \*\*   
## Sodium 0.254813 0.177635 1.434 0.151435   
## CPK -0.348020 0.152638 -2.280 0.022605 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 212.34 on 291 degrees of freedom  
## AIC: 228.34  
##   
## Number of Fisher Scoring iterations: 6

beta2 = coefficients(step.AIC)  
exp (beta2)

## (Intercept) as.factor(EF)1 as.factor(EF)2 `Serum Creatinine`   
## 0.3151904 6.2435301 5.8741537 5.4417878   
## TIME Age Sodium CPK   
## 1.0223878 0.9575269 1.2902209 0.7060845

exp (confint (step.AIC))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.03041315 3.0639360  
## as.factor(EF)1 2.81680717 14.6603964  
## as.factor(EF)2 2.26193946 16.3783840  
## `Serum Creatinine` 2.44722752 12.7054550  
## TIME 1.01656138 1.0290572  
## Age 0.92703754 0.9872768  
## Sodium 0.91337623 1.8388100  
## CPK 0.51876060 0.9463040

step.BIC = step(order1\_fit1, direction = "both", k=log (order1\_fit1$rank + order1\_fit1$df.residual))

## Start: AIC=289.73  
## Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME +   
## Gender + Smoking + BP + Anaemia + Age + Sodium + platelets +   
## CPK  
##   
## Df Deviance AIC  
## - BP 1 209.93 284.03  
## - Anaemia 1 209.94 284.05  
## - Smoking 1 210.03 284.13  
## - Diabetes 1 210.21 284.32  
## - platelets 1 210.40 284.51  
## - Gender 1 211.56 285.67  
## - Sodium 1 212.49 286.59  
## <none> 209.93 289.73  
## - CPK 1 216.35 290.45  
## - Age 1 218.19 292.29  
## - `Serum Creatinine` 1 226.47 300.58  
## - as.factor(EF) 2 234.52 302.93  
## - TIME 1 287.94 362.05  
##   
## Step: AIC=284.03  
## Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME +   
## Gender + Smoking + Anaemia + Age + Sodium + platelets + CPK  
##   
## Df Deviance AIC  
## - Anaemia 1 209.94 278.35  
## - Smoking 1 210.03 278.43  
## - Diabetes 1 210.22 278.62  
## - platelets 1 210.40 278.81  
## - Gender 1 211.59 280.00  
## - Sodium 1 212.49 280.90  
## <none> 209.93 284.03  
## - CPK 1 216.39 284.80  
## - Age 1 218.19 286.59  
## + BP 1 209.93 289.73  
## - `Serum Creatinine` 1 226.64 295.05  
## - as.factor(EF) 2 234.56 297.26  
## - TIME 1 289.64 358.05  
##   
## Step: AIC=278.35  
## Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME +   
## Gender + Smoking + Age + Sodium + platelets + CPK  
##   
## Df Deviance AIC  
## - Smoking 1 210.04 272.74  
## - Diabetes 1 210.24 272.95  
## - platelets 1 210.45 273.15  
## - Gender 1 211.63 274.34  
## - Sodium 1 212.49 275.20  
## <none> 209.94 278.35  
## - CPK 1 216.50 279.21  
## - Age 1 218.32 281.03  
## + Anaemia 1 209.93 284.03  
## + BP 1 209.94 284.05  
## - `Serum Creatinine` 1 226.69 289.40  
## - as.factor(EF) 2 234.69 291.69  
## - TIME 1 291.15 353.86  
##   
## Step: AIC=272.74  
## Survival ~ as.factor(EF) + `Serum Creatinine` + Diabetes + TIME +   
## Gender + Age + Sodium + platelets + CPK  
##   
## Df Deviance AIC  
## - Diabetes 1 210.32 267.33  
## - platelets 1 210.51 267.51  
## - Gender 1 211.73 268.73  
## - Sodium 1 212.54 269.55  
## <none> 210.04 272.74  
## - CPK 1 216.52 273.52  
## - Age 1 218.49 275.49  
## + Smoking 1 209.94 278.35  
## + Anaemia 1 210.03 278.43  
## + BP 1 210.03 278.44  
## - `Serum Creatinine` 1 226.72 283.73  
## - as.factor(EF) 2 234.93 286.23  
## - TIME 1 291.30 348.31  
##   
## Step: AIC=267.33  
## Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Gender +   
## Age + Sodium + platelets + CPK  
##   
## Df Deviance AIC  
## - platelets 1 210.75 262.05  
## - Gender 1 212.14 263.44  
## - Sodium 1 212.96 264.26  
## <none> 210.32 267.33  
## - CPK 1 216.84 268.15  
## - Age 1 218.53 269.83  
## + Diabetes 1 210.04 272.74  
## + Smoking 1 210.24 272.95  
## + Anaemia 1 210.31 273.01  
## + BP 1 210.32 273.03  
## - `Serum Creatinine` 1 227.02 278.32  
## - as.factor(EF) 2 234.97 280.58  
## - TIME 1 291.52 342.83  
##   
## Step: AIC=262.05  
## Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Gender +   
## Age + Sodium + CPK  
##   
## Df Deviance AIC  
## - Gender 1 212.34 257.94  
## - Sodium 1 213.29 258.90  
## <none> 210.75 262.05  
## - CPK 1 216.93 262.53  
## - Age 1 219.29 264.89  
## + platelets 1 210.32 267.33  
## + Diabetes 1 210.51 267.51  
## + Smoking 1 210.69 267.70  
## + Anaemia 1 210.71 267.72  
## + BP 1 210.75 267.75  
## - `Serum Creatinine` 1 227.70 273.31  
## - as.factor(EF) 2 236.02 275.92  
## - TIME 1 291.55 337.15  
##   
## Step: AIC=257.94  
## Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Age +   
## Sodium + CPK  
##   
## Df Deviance AIC  
## - Sodium 1 214.42 254.33  
## - CPK 1 217.80 257.70  
## <none> 212.34 257.94  
## - Age 1 220.17 260.08  
## + Gender 1 210.75 262.05  
## + Diabetes 1 211.98 263.28  
## + platelets 1 212.14 263.44  
## + Anaemia 1 212.23 263.53  
## + Smoking 1 212.23 263.53  
## + BP 1 212.30 263.60  
## - `Serum Creatinine` 1 230.15 270.05  
## - as.factor(EF) 2 236.94 271.14  
## - TIME 1 292.78 332.68  
##   
## Step: AIC=254.33  
## Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Age +   
## CPK  
##   
## Df Deviance AIC  
## <none> 214.42 254.33  
## - CPK 1 220.26 254.46  
## - Age 1 223.31 257.51  
## + Sodium 1 212.34 257.94  
## + Gender 1 213.29 258.90  
## + Diabetes 1 213.95 259.55  
## + platelets 1 214.25 259.85  
## + Smoking 1 214.34 259.94  
## + Anaemia 1 214.39 259.99  
## + BP 1 214.39 259.99  
## - as.factor(EF) 2 238.96 267.47  
## - `Serum Creatinine` 1 235.47 269.67  
## - TIME 1 294.16 328.37

summary(step.BIC)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + `Serum Creatinine` +   
## TIME + Age + CPK, family = binomial)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5558 -0.4809 0.2034 0.5432 2.7908   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.70118 1.11685 -0.628 0.530123   
## as.factor(EF)1 1.81059 0.41494 4.363 1.28e-05 \*\*\*  
## as.factor(EF)2 1.76494 0.49935 3.534 0.000409 \*\*\*  
## `Serum Creatinine` 1.79583 0.41099 4.370 1.24e-05 \*\*\*  
## TIME 0.02171 0.00303 7.165 7.76e-13 \*\*\*  
## Age -0.04575 0.01588 -2.882 0.003948 \*\*   
## CPK -0.35951 0.15270 -2.354 0.018552 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 214.42 on 292 degrees of freedom  
## AIC: 228.42  
##   
## Number of Fisher Scoring iterations: 6

summary(step.AIC)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + `Serum Creatinine` +   
## TIME + Age + Sodium + CPK, family = binomial)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4321 -0.4395 0.1993 0.5316 2.7943   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.154578 1.171306 -0.986 0.324271   
## as.factor(EF)1 1.831546 0.418696 4.374 1.22e-05 \*\*\*  
## as.factor(EF)2 1.770562 0.502666 3.522 0.000428 \*\*\*  
## `Serum Creatinine` 1.694108 0.417850 4.054 5.03e-05 \*\*\*  
## TIME 0.022141 0.003103 7.136 9.59e-13 \*\*\*  
## Age -0.043401 0.015976 -2.717 0.006596 \*\*   
## Sodium 0.254813 0.177635 1.434 0.151435   
## CPK -0.348020 0.152638 -2.280 0.022605 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 212.34 on 291 degrees of freedom  
## AIC: 228.34  
##   
## Number of Fisher Scoring iterations: 6

summary(step.BIC)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + `Serum Creatinine` +   
## TIME + Age + CPK, family = binomial)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5558 -0.4809 0.2034 0.5432 2.7908   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.70118 1.11685 -0.628 0.530123   
## as.factor(EF)1 1.81059 0.41494 4.363 1.28e-05 \*\*\*  
## as.factor(EF)2 1.76494 0.49935 3.534 0.000409 \*\*\*  
## `Serum Creatinine` 1.79583 0.41099 4.370 1.24e-05 \*\*\*  
## TIME 0.02171 0.00303 7.165 7.76e-13 \*\*\*  
## Age -0.04575 0.01588 -2.882 0.003948 \*\*   
## CPK -0.35951 0.15270 -2.354 0.018552 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 214.42 on 292 degrees of freedom  
## AIC: 228.42  
##   
## Number of Fisher Scoring iterations: 6

anova ( step.BIC, step.AIC )

## Analysis of Deviance Table  
##   
## Model 1: Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Age +   
## CPK  
## Model 2: Survival ~ as.factor(EF) + `Serum Creatinine` + TIME + Age +   
## Sodium + CPK  
## Resid. Df Resid. Dev Df Deviance  
## 1 292 214.42   
## 2 291 212.34 1 2.0858

pchisq(2.0858, 1,lower.tail = FALSE)

## [1] 0.1486743

We choose the AIC model to do the interaction effects on becuase it has significant predictors and has a lower residual deviance of 214.42 Residual Deviance on 292 degrees of freedom. The pvalue (0.1486743) which is less than 0.05 the typical cutoff.

# Center quantitative predictors and add interaction effects  
  
my.center = function (y) y - mean (y)  
  
survival$TIME.c = my.center (survival$TIME)  
survival$Age.c = my.center (survival$Age)  
  
  
fit1\_InteractionEffects = glm (Survival ~ (as.factor(EF) + TIME.c + CPK + `Serum Creatinine` + Age.c)^2,  
 data=survival,   
 family = binomial)  
summary (fit1\_InteractionEffects)

##   
## Call:  
## glm(formula = Survival ~ (as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c)^2, family = binomial, data = survival)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.66559 -0.52375 0.06657 0.46041 2.09643   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.270e-01 6.228e-01 -0.846 0.39741   
## as.factor(EF)1 2.848e+00 1.222e+00 2.330 0.01980 \*   
## as.factor(EF)2 2.833e+00 1.456e+00 1.946 0.05167 .   
## TIME.c 1.017e-02 6.982e-03 1.457 0.14509   
## CPK -2.864e-01 3.048e-01 -0.940 0.34730   
## `Serum Creatinine` 1.265e+00 7.936e-01 1.594 0.11083   
## Age.c -6.151e-02 4.715e-02 -1.305 0.19200   
## as.factor(EF)1:TIME.c 2.851e-02 9.996e-03 2.852 0.00434 \*\*  
## as.factor(EF)2:TIME.c 8.913e-03 9.620e-03 0.927 0.35417   
## as.factor(EF)1:CPK -5.273e-01 4.297e-01 -1.227 0.21977   
## as.factor(EF)2:CPK -6.702e-01 4.528e-01 -1.480 0.13887   
## as.factor(EF)1:`Serum Creatinine` 1.380e+00 1.021e+00 1.352 0.17640   
## as.factor(EF)2:`Serum Creatinine` 3.330e-01 1.310e+00 0.254 0.79931   
## as.factor(EF)1:Age.c -4.433e-02 4.349e-02 -1.019 0.30805   
## as.factor(EF)2:Age.c -3.702e-03 4.482e-02 -0.083 0.93418   
## TIME.c:CPK -2.215e-03 2.796e-03 -0.792 0.42824   
## TIME.c:`Serum Creatinine` 1.174e-02 6.719e-03 1.747 0.08069 .   
## TIME.c:Age.c -7.281e-05 3.131e-04 -0.233 0.81611   
## CPK:`Serum Creatinine` 1.469e-01 3.668e-01 0.401 0.68874   
## CPK:Age.c 1.972e-02 1.636e-02 1.205 0.22821   
## `Serum Creatinine`:Age.c -2.237e-02 3.951e-02 -0.566 0.57137   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 194.42 on 278 degrees of freedom  
## AIC: 236.42  
##   
## Number of Fisher Scoring iterations: 7

# Final Stepwise Regression

step.AIC\_final = step (fit1\_InteractionEffects, direction = "both")

## Start: AIC=236.42  
## Survival ~ (as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c)^2  
##   
## Df Deviance AIC  
## - as.factor(EF):Age.c 2 195.65 233.65  
## - as.factor(EF):`Serum Creatinine` 2 196.35 234.35  
## - TIME.c:Age.c 1 194.47 234.47  
## - CPK:`Serum Creatinine` 1 194.58 234.58  
## - `Serum Creatinine`:Age.c 1 194.74 234.74  
## - TIME.c:CPK 1 195.05 235.05  
## - as.factor(EF):CPK 2 197.13 235.13  
## - CPK:Age.c 1 195.94 235.94  
## <none> 194.42 236.42  
## - TIME.c:`Serum Creatinine` 1 197.40 237.40  
## - as.factor(EF):TIME.c 2 205.48 243.48  
##   
## Step: AIC=233.65  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + as.factor(EF):CPK + as.factor(EF):`Serum Creatinine` +   
## TIME.c:CPK + TIME.c:`Serum Creatinine` + TIME.c:Age.c + CPK:`Serum Creatinine` +   
## CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - TIME.c:Age.c 1 195.66 231.66  
## - CPK:`Serum Creatinine` 1 195.70 231.70  
## - `Serum Creatinine`:Age.c 1 196.00 232.00  
## - as.factor(EF):`Serum Creatinine` 2 198.10 232.10  
## - as.factor(EF):CPK 2 198.34 232.34  
## - TIME.c:CPK 1 196.41 232.41  
## - CPK:Age.c 1 196.80 232.80  
## <none> 195.65 233.65  
## - TIME.c:`Serum Creatinine` 1 199.21 235.21  
## + as.factor(EF):Age.c 2 194.42 236.42  
## - as.factor(EF):TIME.c 2 206.16 240.16  
##   
## Step: AIC=231.66  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + as.factor(EF):CPK + as.factor(EF):`Serum Creatinine` +   
## TIME.c:CPK + TIME.c:`Serum Creatinine` + CPK:`Serum Creatinine` +   
## CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - CPK:`Serum Creatinine` 1 195.71 229.71  
## - `Serum Creatinine`:Age.c 1 196.00 230.00  
## - as.factor(EF):`Serum Creatinine` 2 198.17 230.17  
## - as.factor(EF):CPK 2 198.39 230.39  
## - TIME.c:CPK 1 196.41 230.41  
## - CPK:Age.c 1 196.81 230.81  
## <none> 195.66 231.66  
## - TIME.c:`Serum Creatinine` 1 199.22 233.22  
## + TIME.c:Age.c 1 195.65 233.65  
## + as.factor(EF):Age.c 2 194.47 234.47  
## - as.factor(EF):TIME.c 2 206.66 238.66  
##   
## Step: AIC=229.71  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + as.factor(EF):CPK + as.factor(EF):`Serum Creatinine` +   
## TIME.c:CPK + TIME.c:`Serum Creatinine` + CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - `Serum Creatinine`:Age.c 1 196.08 228.08  
## - as.factor(EF):CPK 2 198.39 228.39  
## - as.factor(EF):`Serum Creatinine` 2 198.42 228.42  
## - TIME.c:CPK 1 196.56 228.56  
## - CPK:Age.c 1 196.81 228.81  
## <none> 195.71 229.71  
## - TIME.c:`Serum Creatinine` 1 199.46 231.46  
## + CPK:`Serum Creatinine` 1 195.66 231.66  
## + TIME.c:Age.c 1 195.70 231.70  
## + as.factor(EF):Age.c 2 194.62 232.62  
## - as.factor(EF):TIME.c 2 206.68 236.68  
##   
## Step: AIC=228.08  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + as.factor(EF):CPK + as.factor(EF):`Serum Creatinine` +   
## TIME.c:CPK + TIME.c:`Serum Creatinine` + CPK:Age.c  
##   
## Df Deviance AIC  
## - as.factor(EF):`Serum Creatinine` 2 198.56 226.56  
## - TIME.c:CPK 1 197.06 227.06  
## - CPK:Age.c 1 197.11 227.11  
## - as.factor(EF):CPK 2 199.24 227.24  
## <none> 196.08 228.08  
## + `Serum Creatinine`:Age.c 1 195.71 229.71  
## + CPK:`Serum Creatinine` 1 196.00 230.00  
## - TIME.c:`Serum Creatinine` 1 200.02 230.02  
## + TIME.c:Age.c 1 196.08 230.08  
## + as.factor(EF):Age.c 2 194.96 230.96  
## - as.factor(EF):TIME.c 2 207.93 235.93  
##   
## Step: AIC=226.56  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + as.factor(EF):CPK + TIME.c:CPK +   
## TIME.c:`Serum Creatinine` + CPK:Age.c  
##   
## Df Deviance AIC  
## - as.factor(EF):CPK 2 201.53 225.53  
## - TIME.c:CPK 1 199.75 225.75  
## - CPK:Age.c 1 200.00 226.00  
## <none> 198.56 226.56  
## - TIME.c:`Serum Creatinine` 1 201.49 227.49  
## + as.factor(EF):`Serum Creatinine` 2 196.08 228.08  
## + CPK:`Serum Creatinine` 1 198.28 228.28  
## + `Serum Creatinine`:Age.c 1 198.42 228.42  
## + TIME.c:Age.c 1 198.52 228.52  
## + as.factor(EF):Age.c 2 197.14 229.14  
## - as.factor(EF):TIME.c 2 208.44 232.44  
##   
## Step: AIC=225.53  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:CPK + TIME.c:`Serum Creatinine` +   
## CPK:Age.c  
##   
## Df Deviance AIC  
## - TIME.c:CPK 1 201.94 223.94  
## - CPK:Age.c 1 201.98 223.98  
## <none> 201.53 225.53  
## - TIME.c:`Serum Creatinine` 1 204.46 226.46  
## + as.factor(EF):CPK 2 198.56 226.56  
## + `Serum Creatinine`:Age.c 1 201.14 227.14  
## + as.factor(EF):`Serum Creatinine` 2 199.24 227.24  
## + CPK:`Serum Creatinine` 1 201.44 227.44  
## + TIME.c:Age.c 1 201.50 227.50  
## + as.factor(EF):Age.c 2 200.11 228.11  
## - as.factor(EF):TIME.c 2 210.57 230.57  
##   
## Step: AIC=223.94  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine` +   
## CPK:Age.c  
##   
## Df Deviance AIC  
## - CPK:Age.c 1 202.47 222.47  
## <none> 201.94 223.94  
## - TIME.c:`Serum Creatinine` 1 204.79 224.79  
## + as.factor(EF):`Serum Creatinine` 2 199.50 225.50  
## + `Serum Creatinine`:Age.c 1 201.52 225.52  
## + TIME.c:CPK 1 201.53 225.53  
## + CPK:`Serum Creatinine` 1 201.74 225.74  
## + as.factor(EF):CPK 2 199.75 225.75  
## + TIME.c:Age.c 1 201.94 225.94  
## + as.factor(EF):Age.c 2 200.44 226.44  
## - as.factor(EF):TIME.c 2 210.65 228.65  
##   
## Step: AIC=222.47  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine`  
##   
## Df Deviance AIC  
## <none> 202.47 222.47  
## - TIME.c:`Serum Creatinine` 1 205.25 223.25  
## + as.factor(EF):`Serum Creatinine` 2 199.77 223.77  
## + CPK:Age.c 1 201.94 223.94  
## + TIME.c:CPK 1 201.98 223.98  
## + `Serum Creatinine`:Age.c 1 202.20 224.20  
## + CPK:`Serum Creatinine` 1 202.31 224.31  
## + TIME.c:Age.c 1 202.46 224.46  
## + as.factor(EF):CPK 2 201.14 225.14  
## + as.factor(EF):Age.c 2 201.19 225.19  
## - CPK 1 209.49 227.49  
## - as.factor(EF):TIME.c 2 211.91 227.91  
## - Age.c 1 212.49 230.49

summary(step.AIC\_final)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine`,   
## family = binomial, data = survival)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.71103 -0.49728 0.08546 0.50304 2.33555   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.683582 0.423220 -1.615 0.10627   
## as.factor(EF)1 2.555116 0.564167 4.529 5.93e-06 \*\*\*  
## as.factor(EF)2 1.955026 0.694497 2.815 0.00488 \*\*   
## TIME.c 0.008115 0.005224 1.553 0.12032   
## CPK -0.401538 0.156240 -2.570 0.01017 \*   
## `Serum Creatinine` 1.924515 0.443882 4.336 1.45e-05 \*\*\*  
## Age.c -0.050382 0.016646 -3.027 0.00247 \*\*   
## as.factor(EF)1:TIME.c 0.022274 0.008176 2.724 0.00644 \*\*   
## as.factor(EF)2:TIME.c 0.008078 0.009461 0.854 0.39321   
## TIME.c:`Serum Creatinine` 0.010449 0.006271 1.666 0.09565 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 202.47 on 289 degrees of freedom  
## AIC: 222.47  
##   
## Number of Fisher Scoring iterations: 7

step.BIC\_final = step (fit1\_InteractionEffects, direction = "both",k=log (fit1\_InteractionEffects$rank + fit1\_InteractionEffects$df.residual))

## Start: AIC=314.13  
## Survival ~ (as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c)^2  
##   
## Df Deviance AIC  
## - as.factor(EF):Age.c 2 195.65 303.96  
## - as.factor(EF):`Serum Creatinine` 2 196.35 304.65  
## - as.factor(EF):CPK 2 197.13 305.44  
## - TIME.c:Age.c 1 194.47 308.48  
## - CPK:`Serum Creatinine` 1 194.58 308.59  
## - `Serum Creatinine`:Age.c 1 194.74 308.74  
## - TIME.c:CPK 1 195.05 309.06  
## - CPK:Age.c 1 195.94 309.94  
## - TIME.c:`Serum Creatinine` 1 197.40 311.41  
## - as.factor(EF):TIME.c 2 205.48 313.79  
## <none> 194.42 314.13  
##   
## Step: AIC=303.96  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + as.factor(EF):CPK + as.factor(EF):`Serum Creatinine` +   
## TIME.c:CPK + TIME.c:`Serum Creatinine` + TIME.c:Age.c + CPK:`Serum Creatinine` +   
## CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - as.factor(EF):`Serum Creatinine` 2 198.10 295.01  
## - as.factor(EF):CPK 2 198.34 295.25  
## - TIME.c:Age.c 1 195.66 298.27  
## - CPK:`Serum Creatinine` 1 195.70 298.31  
## - `Serum Creatinine`:Age.c 1 196.00 298.61  
## - TIME.c:CPK 1 196.41 299.02  
## - CPK:Age.c 1 196.80 299.41  
## - TIME.c:`Serum Creatinine` 1 199.21 301.81  
## - as.factor(EF):TIME.c 2 206.16 303.07  
## <none> 195.65 303.96  
## + as.factor(EF):Age.c 2 194.42 314.13  
##   
## Step: AIC=295.01  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + as.factor(EF):CPK + TIME.c:CPK +   
## TIME.c:`Serum Creatinine` + TIME.c:Age.c + CPK:`Serum Creatinine` +   
## CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - as.factor(EF):CPK 2 200.98 286.49  
## - TIME.c:Age.c 1 198.17 289.38  
## - `Serum Creatinine`:Age.c 1 198.24 289.45  
## - CPK:`Serum Creatinine` 1 198.35 289.56  
## - TIME.c:CPK 1 199.04 290.24  
## - CPK:Age.c 1 199.86 291.07  
## - TIME.c:`Serum Creatinine` 1 200.57 291.78  
## - as.factor(EF):TIME.c 2 206.98 292.48  
## <none> 198.10 295.01  
## + as.factor(EF):`Serum Creatinine` 2 195.65 303.96  
## + as.factor(EF):Age.c 2 196.35 304.65  
##   
## Step: AIC=286.49  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:CPK + TIME.c:`Serum Creatinine` +   
## TIME.c:Age.c + CPK:`Serum Creatinine` + CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - CPK:`Serum Creatinine` 1 201.04 280.85  
## - TIME.c:Age.c 1 201.07 280.88  
## - TIME.c:CPK 1 201.34 281.14  
## - `Serum Creatinine`:Age.c 1 201.41 281.21  
## - CPK:Age.c 1 201.66 281.47  
## - as.factor(EF):TIME.c 2 209.02 283.12  
## - TIME.c:`Serum Creatinine` 1 203.44 283.25  
## <none> 200.98 286.49  
## + as.factor(EF):CPK 2 198.10 295.01  
## + as.factor(EF):`Serum Creatinine` 2 198.34 295.25  
## + as.factor(EF):Age.c 2 199.28 296.18  
##   
## Step: AIC=280.85  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:CPK + TIME.c:`Serum Creatinine` +   
## TIME.c:Age.c + CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - TIME.c:Age.c 1 201.14 275.24  
## - TIME.c:CPK 1 201.49 275.60  
## - `Serum Creatinine`:Age.c 1 201.50 275.61  
## - CPK:Age.c 1 201.69 275.80  
## - as.factor(EF):TIME.c 2 209.03 277.44  
## - TIME.c:`Serum Creatinine` 1 203.68 277.79  
## <none> 201.04 280.85  
## + CPK:`Serum Creatinine` 1 200.98 286.49  
## + as.factor(EF):`Serum Creatinine` 2 198.35 289.56  
## + as.factor(EF):CPK 2 198.35 289.56  
## + as.factor(EF):Age.c 2 199.41 290.62  
##   
## Step: AIC=275.24  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:CPK + TIME.c:`Serum Creatinine` +   
## CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - TIME.c:CPK 1 201.52 269.93  
## - `Serum Creatinine`:Age.c 1 201.53 269.94  
## - CPK:Age.c 1 201.71 270.12  
## - as.factor(EF):TIME.c 2 209.40 272.11  
## - TIME.c:`Serum Creatinine` 1 203.71 272.11  
## <none> 201.14 275.24  
## + TIME.c:Age.c 1 201.04 280.85  
## + CPK:`Serum Creatinine` 1 201.07 280.88  
## + as.factor(EF):`Serum Creatinine` 2 198.39 283.90  
## + as.factor(EF):CPK 2 198.42 283.93  
## + as.factor(EF):Age.c 2 199.58 285.09  
##   
## Step: AIC=269.93  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine` +   
## CPK:Age.c + `Serum Creatinine`:Age.c  
##   
## Df Deviance AIC  
## - `Serum Creatinine`:Age.c 1 201.94 264.65  
## - CPK:Age.c 1 202.20 264.90  
## - as.factor(EF):TIME.c 2 209.48 266.48  
## - TIME.c:`Serum Creatinine` 1 204.00 266.71  
## <none> 201.52 269.93  
## + TIME.c:CPK 1 201.14 275.24  
## + CPK:`Serum Creatinine` 1 201.37 275.47  
## + TIME.c:Age.c 1 201.49 275.60  
## + as.factor(EF):`Serum Creatinine` 2 198.61 278.42  
## + as.factor(EF):CPK 2 199.53 279.34  
## + as.factor(EF):Age.c 2 199.90 279.70  
##   
## Step: AIC=264.65  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine` +   
## CPK:Age.c  
##   
## Df Deviance AIC  
## - CPK:Age.c 1 202.47 259.48  
## - TIME.c:`Serum Creatinine` 1 204.79 261.80  
## - as.factor(EF):TIME.c 2 210.65 261.95  
## <none> 201.94 264.65  
## + `Serum Creatinine`:Age.c 1 201.52 269.93  
## + TIME.c:CPK 1 201.53 269.94  
## + CPK:`Serum Creatinine` 1 201.74 270.14  
## + TIME.c:Age.c 1 201.94 270.35  
## + as.factor(EF):`Serum Creatinine` 2 199.50 273.60  
## + as.factor(EF):CPK 2 199.75 273.85  
## + as.factor(EF):Age.c 2 200.44 274.54  
##   
## Step: AIC=259.48  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine`  
##   
## Df Deviance AIC  
## - TIME.c:`Serum Creatinine` 1 205.25 256.55  
## - as.factor(EF):TIME.c 2 211.91 257.51  
## <none> 202.47 259.48  
## - CPK 1 209.49 260.79  
## - Age.c 1 212.49 263.80  
## + CPK:Age.c 1 201.94 264.65  
## + TIME.c:CPK 1 201.98 264.68  
## + `Serum Creatinine`:Age.c 1 202.20 264.90  
## + CPK:`Serum Creatinine` 1 202.31 265.02  
## + TIME.c:Age.c 1 202.46 265.17  
## + as.factor(EF):`Serum Creatinine` 2 199.77 268.18  
## + as.factor(EF):CPK 2 201.14 269.55  
## + as.factor(EF):Age.c 2 201.19 269.59  
##   
## Step: AIC=256.55  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c  
##   
## Df Deviance AIC  
## - as.factor(EF):TIME.c 2 214.42 254.33  
## <none> 205.25 256.55  
## - CPK 1 212.00 257.60  
## + TIME.c:`Serum Creatinine` 1 202.47 259.48  
## - Age.c 1 214.56 260.16  
## + `Serum Creatinine`:Age.c 1 204.61 261.62  
## + CPK:`Serum Creatinine` 1 204.79 261.80  
## + CPK:Age.c 1 204.79 261.80  
## + TIME.c:CPK 1 204.84 261.85  
## + TIME.c:Age.c 1 205.13 262.13  
## + as.factor(EF):`Serum Creatinine` 2 203.64 266.34  
## + as.factor(EF):Age.c 2 203.80 266.51  
## + as.factor(EF):CPK 2 203.90 266.61  
## - `Serum Creatinine` 1 222.60 268.21  
##   
## Step: AIC=254.33  
## Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c  
##   
## Df Deviance AIC  
## <none> 214.42 254.33  
## - CPK 1 220.26 254.46  
## + as.factor(EF):TIME.c 2 205.25 256.55  
## - Age.c 1 223.31 257.51  
## + TIME.c:`Serum Creatinine` 1 211.91 257.51  
## + TIME.c:Age.c 1 212.87 258.47  
## + `Serum Creatinine`:Age.c 1 212.96 258.56  
## + CPK:Age.c 1 213.32 258.92  
## + CPK:`Serum Creatinine` 1 214.18 259.78  
## + TIME.c:CPK 1 214.27 259.87  
## + as.factor(EF):CPK 2 213.41 264.72  
## + as.factor(EF):Age.c 2 213.45 264.75  
## + as.factor(EF):`Serum Creatinine` 2 213.72 265.02  
## - as.factor(EF) 2 238.96 267.47  
## - `Serum Creatinine` 1 235.47 269.67  
## - TIME.c 1 294.16 328.37

summary(step.BIC\_final)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c, family = binomial, data = survival)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5558 -0.4809 0.2034 0.5432 2.7908   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.65682 0.44189 -1.486 0.137179   
## as.factor(EF)1 1.81059 0.41494 4.363 1.28e-05 \*\*\*  
## as.factor(EF)2 1.76494 0.49935 3.534 0.000409 \*\*\*  
## TIME.c 0.02171 0.00303 7.165 7.76e-13 \*\*\*  
## CPK -0.35951 0.15270 -2.354 0.018552 \*   
## `Serum Creatinine` 1.79583 0.41099 4.370 1.24e-05 \*\*\*  
## Age.c -0.04575 0.01588 -2.882 0.003948 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 214.42 on 292 degrees of freedom  
## AIC: 228.42  
##   
## Number of Fisher Scoring iterations: 6

summary(step.AIC\_final)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine`,   
## family = binomial, data = survival)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.71103 -0.49728 0.08546 0.50304 2.33555   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.683582 0.423220 -1.615 0.10627   
## as.factor(EF)1 2.555116 0.564167 4.529 5.93e-06 \*\*\*  
## as.factor(EF)2 1.955026 0.694497 2.815 0.00488 \*\*   
## TIME.c 0.008115 0.005224 1.553 0.12032   
## CPK -0.401538 0.156240 -2.570 0.01017 \*   
## `Serum Creatinine` 1.924515 0.443882 4.336 1.45e-05 \*\*\*  
## Age.c -0.050382 0.016646 -3.027 0.00247 \*\*   
## as.factor(EF)1:TIME.c 0.022274 0.008176 2.724 0.00644 \*\*   
## as.factor(EF)2:TIME.c 0.008078 0.009461 0.854 0.39321   
## TIME.c:`Serum Creatinine` 0.010449 0.006271 1.666 0.09565 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 202.47 on 289 degrees of freedom  
## AIC: 222.47  
##   
## Number of Fisher Scoring iterations: 7

summary(step.BIC\_final)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c, family = binomial, data = survival)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5558 -0.4809 0.2034 0.5432 2.7908   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.65682 0.44189 -1.486 0.137179   
## as.factor(EF)1 1.81059 0.41494 4.363 1.28e-05 \*\*\*  
## as.factor(EF)2 1.76494 0.49935 3.534 0.000409 \*\*\*  
## TIME.c 0.02171 0.00303 7.165 7.76e-13 \*\*\*  
## CPK -0.35951 0.15270 -2.354 0.018552 \*   
## `Serum Creatinine` 1.79583 0.41099 4.370 1.24e-05 \*\*\*  
## Age.c -0.04575 0.01588 -2.882 0.003948 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 214.42 on 292 degrees of freedom  
## AIC: 228.42  
##   
## Number of Fisher Scoring iterations: 6

anova (step.AIC\_final, step.BIC\_final)

## Analysis of Deviance Table  
##   
## Model 1: Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine`  
## Model 2: Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c  
## Resid. Df Resid. Dev Df Deviance  
## 1 289 202.47   
## 2 292 214.42 -3 -11.949

pchisq(11.949, 3, lower.tail = FALSE)

## [1] 0.007559923

We choose AIC model as final becuase it's Residual Deviance is 202.47 which less than the Residual Deviance 214.42 of the BIC model. The pvalue (0.0075599) is less than 0.05 which means that the two models are statistically different from each other.

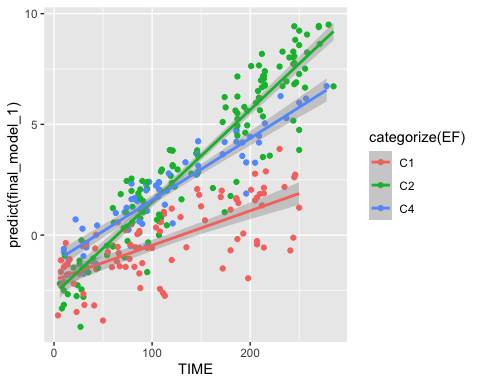
# Final Model:

final\_model\_1 = step.AIC\_final  
summary(final\_model\_1)

##   
## Call:  
## glm(formula = Survival ~ as.factor(EF) + TIME.c + CPK + `Serum Creatinine` +   
## Age.c + as.factor(EF):TIME.c + TIME.c:`Serum Creatinine`,   
## family = binomial, data = survival)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.71103 -0.49728 0.08546 0.50304 2.33555   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.683582 0.423220 -1.615 0.10627   
## as.factor(EF)1 2.555116 0.564167 4.529 5.93e-06 \*\*\*  
## as.factor(EF)2 1.955026 0.694497 2.815 0.00488 \*\*   
## TIME.c 0.008115 0.005224 1.553 0.12032   
## CPK -0.401538 0.156240 -2.570 0.01017 \*   
## `Serum Creatinine` 1.924515 0.443882 4.336 1.45e-05 \*\*\*  
## Age.c -0.050382 0.016646 -3.027 0.00247 \*\*   
## as.factor(EF)1:TIME.c 0.022274 0.008176 2.724 0.00644 \*\*   
## as.factor(EF)2:TIME.c 0.008078 0.009461 0.854 0.39321   
## TIME.c:`Serum Creatinine` 0.010449 0.006271 1.666 0.09565 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 375.35 on 298 degrees of freedom  
## Residual deviance: 202.47 on 289 degrees of freedom  
## AIC: 222.47  
##   
## Number of Fisher Scoring iterations: 7

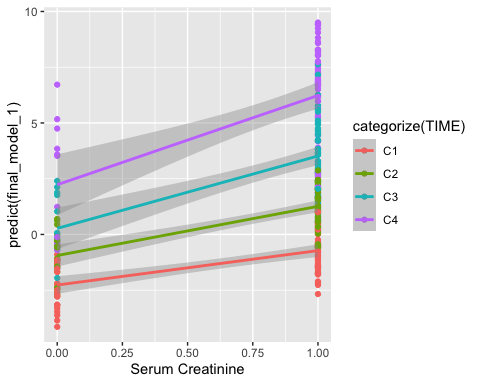
### Interaction plots  
  
# New categorize function  
  
categorize = function (x, quantiles=(1:3)/4) {  
 cutoffs = quantile (x, quantiles)  
 n.cutoffs = length (cutoffs)  
 result = rep ("C1", length (x))  
 for (j in 1:n.cutoffs) {  
 result [x > cutoffs [j]] = paste ("C", j+1, sep="")  
 }  
 return (result)  
}  
  
library (ggplot2)  
  
qplot (TIME, predict (final\_model\_1), data=survival, color=categorize (EF)) + geom\_smooth (method="lm")

## `geom\_smooth()` using formula 'y ~ x'



qplot (`Serum Creatinine`, predict (final\_model\_1), data=survival, color=categorize (TIME)) + geom\_smooth (method="lm")

## `geom\_smooth()` using formula 'y ~ x'



The interaction plot between time and Ejection Fraction shows that the predicted logits of surviving cardio-vascular disease increases over time for all ejection fraction (EF) categories(EF<30 (coded as 0), 301.5(indicator of renal dysfunction coded as 0) and < 1.5(normal level coded as 1). We can see from the positive slope of all four of the TIME graphs that all patients with renal dysfunction (0) have lower predicted logits of surviving than those who do not (1). All patients with follow up time in C4 (75 quartile),regardless of whether they have renal dysfunction, have higher predicted logits of surviving cardio-vascular disease than all other patients in C1,2 & 3, regardless of whether those patients do not have renal dysfunction. Patients with follow up time in C4 (75 quartile), also have the most noticeable change in the predicted logits for surviving cardio-vascular disease between patients with renal dysfunction(0) and patients without (0).

exp (confint (final\_model\_1))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.2156857 1.1469405  
## as.factor(EF)1 4.6119614 43.0804413  
## as.factor(EF)2 2.0696061 33.5811069  
## TIME.c 0.9980162 1.0189091  
## CPK 0.4877742 0.9025785  
## `Serum Creatinine` 2.9247387 16.8167300  
## Age.c 0.9191504 0.9814366  
## as.factor(EF)1:TIME.c 1.0075532 1.0408464  
## as.factor(EF)2:TIME.c 0.9911966 1.0294466  
## TIME.c:`Serum Creatinine` 0.9981325 1.0231982

exp (final\_model\_1$coefficients)

## (Intercept) as.factor(EF)1 as.factor(EF)2   
## 0.5048058 12.8727899 7.0641014   
## TIME.c CPK `Serum Creatinine`   
## 1.0081478 0.6692896 6.8518241   
## Age.c as.factor(EF)1:TIME.c as.factor(EF)2:TIME.c   
## 0.9508661 1.0225241 1.0081103   
## TIME.c:`Serum Creatinine`   
## 1.0105039

An increase in a unit of Creatinine Phosphokinase (CPK) decreases a patients of odds of surviving by 0.331-folds. A unit increase in Age decreases a patients odds of surviving by a range of 0.0808 to 0.01856-folds.

preds = predict (final\_model\_1, se.fit = T)   
pred.df = cbind.data.frame (survival, as.data.frame (preds))   
pred.df$lwr = pred.df$fit - 1.96 \* pred.df$se.fit   
pred.df$upr = pred.df$fit + 1.96 \* pred.df$se.fit   
pred.df$fit.pr = round (exp (pred.df$fit) / (1 + exp (pred.df$fit)), 3)   
pred.df$lwr.pr = round (exp (pred.df$lwr) / (1 + exp (pred.df$lwr)), 3)   
pred.df$upr.pr = round (exp (pred.df$upr) / (1 + exp (pred.df$upr)), 3)  
head(pred.df)

## TIME Survival Gender Smoking Diabetes BP Anaemia Age EF Sodium  
## 1 97 1 0 0 0 0 1 43 2 1  
## 2 180 1 1 1 1 0 1 73 0 3  
## 3 31 0 1 1 0 1 0 70 0 0  
## 4 87 1 1 0 0 0 1 65 0 3  
## 5 113 1 1 0 0 0 0 64 2 1  
## 6 10 0 1 0 0 0 1 75 0 1  
## Serum Creatinine platelets CPK index Residual leverage rstudent TIME.c  
## 1 1 1 2 1 1.1452 0.0500 0.5280 -33.26087  
## 2 1 0 1 2 1.1938 0.0733 0.6080 49.73913  
## 3 0 2 3 3 -1.0117 0.0071 -0.1526 -99.26087  
## 4 1 2 2 4 1.9198 0.0922 1.1824 -43.26087  
## 5 1 1 3 5 1.2178 0.0479 0.6364 -17.26087  
## 6 1 0 1 6 -1.0803 0.0286 -0.3960 -120.26087  
## Age.c fit se.fit residual.scale lwr upr fit.pr  
## 1 -17.833893 2.4052705 0.5758996 1 1.27650739 3.53403365 0.917  
## 2 12.166107 1.1497924 0.5501004 1 0.07159569 2.22798910 0.759  
## 3 9.166107 -3.1554807 0.7087235 1 -4.54457862 -1.76638269 0.041  
## 4 4.166107 -0.5751296 0.3313614 1 -1.22459784 0.07433868 0.360  
## 5 3.166107 1.3719734 0.5669334 1 0.26078397 2.48316288 0.798  
## 6 14.166107 -2.1068292 0.5736153 1 -3.23111513 -0.98254321 0.108  
## lwr.pr upr.pr  
## 1 0.782 0.972  
## 2 0.518 0.903  
## 3 0.011 0.146  
## 4 0.227 0.519  
## 5 0.565 0.923  
## 6 0.038 0.272

# Patients Predictions:

# Predictions for patients in their 70s and older.   
pred.df[c(2, 17, 88, 48, 40),c(1,2,3,8,17:20)]

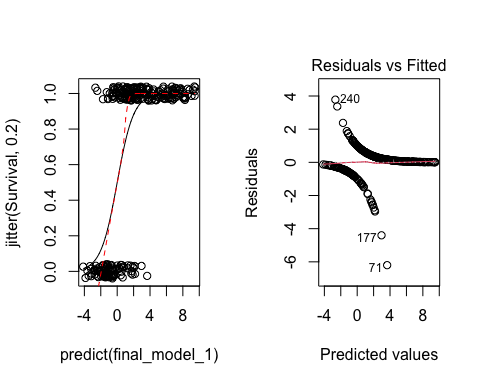
## TIME Survival Gender Age rstudent TIME.c Age.c fit  
## 2 180 1 1 73 0.6080 49.7391304 12.16611 1.1497924  
## 17 28 0 1 85 -0.5314 -102.2608696 24.16611 -2.4535981  
## 88 130 0 0 80 -1.3707 -0.2608696 19.16611 1.6151530  
## 48 109 0 1 80 -1.4417 -21.2608696 19.16611 0.7575546  
## 40 24 0 0 95 -0.7606 -106.2608696 34.16611 -2.2647920

# Predictions for patients in their 40s and 50s   
pred.df[c(103, 66, 83, 21, 11),c(1,2,3,8,17:20)]

## TIME Survival Gender Age rstudent TIME.c Age.c fit  
## 103 61 0 1 45 -2.1708 -69.26087 -15.833893 0.5607006  
## 66 11 0 1 45 -0.5346 -119.26087 -15.833893 -1.3798815  
## 83 196 0 0 54 -2.1218 65.73913 -6.833893 1.8771453  
## 21 192 1 1 50 0.1934 61.73913 -10.833893 4.1820001  
## 11 119 1 1 50 0.4907 -11.26087 -10.833893 2.6773956

# Model Diagnostics:

# Diagnostic plots for model hp9  
  
par (mfrow=c(1,2))  
plot (jitter (Survival, 0.2) ~ predict (final\_model\_1), data=survival)  
logit.9 = predict (final\_model\_1)  
ord9 = order (logit.9)  
pihat.9 = exp (logit.9) / (1 + exp (logit.9))  
lines (logit.9 [ord9], pihat.9 [ord9])  
  
# Lowess fit  
lines (lowess (logit.9, survival$Survival), col='red', lty=2)  
plot (final\_model\_1, which=1)



# Likelihood ratio test for first model  
  
pchisq (order1\_fit1$null.deviance - order1\_fit1$deviance,  
 order1\_fit1$df.null - order1\_fit1$df.residual, lower.tail = F)

## [1] 1.570568e-28

The pvalue (1.570568e-28) is less than 0.05, therefore we reject the null hypothesis and conclude that at least one of our parameters is different from zero.

# Likelihood ratio test for final model  
  
pchisq (final\_model\_1$null.deviance - final\_model\_1$deviance,  
 final\_model\_1$df.null - final\_model\_1$df.residual, lower.tail = F)

## [1] 1.553978e-32

The pvalue (1.553978e-32) is less than 0.05, therefore we reject the null hypothesis and conclude that at least one of our parameters is different from zero.

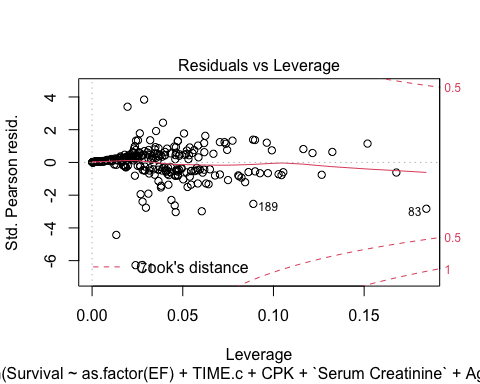
# Goodness of fit test  
  
pchisq (final\_model\_1$deviance, final\_model\_1$df.residual, lower.tail = F)

## [1] 0.9999696

The p\_value(0.9999696) is bigger than (0.05), suggesting that there is no significant lack of fit in the model.

# Influence diagnostics analysis:

plot(final\_model\_1,which = 5)



# ROC plot:

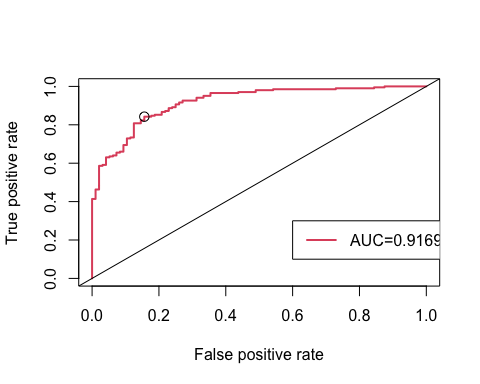
# ROC Curve for model final\_model\_1  
  
# ROC curve - install package ROCR  
par (mfrow=c(1,1))  
library(ROCR)  
pred1 <- prediction(final\_model\_1$fitted.values, final\_model\_1$y)  
perf1 <- performance(pred1,"tpr","fpr")  
auc1 <- performance(pred1,"auc")@y.values[[1]]  
auc1

## [1] 0.9169232

plot(perf1, lwd=2, col=2)  
abline(0,1)  
legend(0.6, 0.3, c(paste ("AUC=", round (auc1, 4), sep="")), lwd=2, col=2)  
  
# Extract the X and Y values from the ROC plot, as well as the probability cutoffs  
roc.x = slot (perf1, "x.values") [[1]]  
roc.y = slot (perf1, "y.values") [[1]]  
cutoffs = slot (perf1, "alpha.values") [[1]]  
  
auc.table = cbind.data.frame(cutoff=pred1@cutoffs,   
 tp=pred1@tp, fp=pred1@fp, tn=pred1@tn, fn=pred1@fn)  
names (auc.table) = c("Cutoff", "TP", "FP", "TN", "FN")  
auc.table$sensitivity = auc.table$TP / (auc.table$TP + auc.table$FN)  
auc.table$specificity = auc.table$TN / (auc.table$TN + auc.table$FP)  
auc.table$FalsePosRate = 1 - auc.table$specificity  
auc.table$sens\_spec = auc.table$sensitivity + auc.table$specificity  
  
# Find the row(s) in the AUC table where sensitivity + specificity is maximized  
  
auc.best = auc.table [auc.table$sens\_spec == max (auc.table$sens\_spec),]  
auc.best

## Cutoff TP FP TN FN sensitivity specificity FalsePosRate sens\_spec  
## 298 0.6412728 171 15 81 32 0.8423645 0.84375 0.15625 1.686115

# Plot the maximum point(s) on the ROC plot  
  
points (auc.best$FalsePosRate, auc.best$sensitivity, cex=1.3)



The AUC is 0.916, meaning that the model correctly predicted for 91.6% of the cases in our data set. This suggests that our final model predicts surviving cardio-vascular disease better than simply guessing. Our cutoff for determining predictions of failure or no failure was 0.6412. This has a false positive rate of 156 and a sensitivity (probability of a true positive prediction) of 0.84.

# Final Conclusions:

We can conclude that, among the predictor variables from our data set, the presence of Ejection Fraction, Time, Creatinine Phosphokinase (CPK), Serum Creatinine and Age were helpful in explaining patients' survival Cardiovascular Heart Disease (CHD). We also found a significant interaction effect between Time and Ejection Fraction. The AUC of this model is 0.916, showing that these five variables and the four interaction plots help predict patients' survival Cardiovascular Heart Disease (CHD) better than simply guessing, and the low p-value of 1.553978e-32 from the Likelihood Ratio test confirmed the significance of the model. The model’s sensitivity was 0.84 and specificity was 0.84375, suggesting that the model does a much better job predicting non-failures than failures. This analysis begs the question of what other variables can explain dam failure, as the variables from this set could not fully account for dam failure.

Some follow-up questions to this analysis could include:

1. Our analysis shows that an increase in age decreases a patient's chance of survival; it also indicates that if a patient is older than 75 years, with EF of 1 or 2, no renal dysfunction and has a follow-up time after 100 days(not sick). Those patients will have a better survival rate than those who have the same age but have renal dysfunction and EF of 0. Will creating an educational program that prevents renal dysfunction and maintains an EF level of 1 or 2 to improve the survival rate for patients older than 75 years and who have a follow-up time after 100 days.
2. Are there other factors that could be measured, such as obesity, cholesterol, family history and physical inactivity that could improve the predictability of the model?