DSC 424 – Advanced Data Analysis Final Project

Patient Survival Prediction Analysis Appendix

Shanele Youngquist, Mary Doerries, Cristian Casado, Can Dai, Michael Goss
DePaul University

Appendix

Appendix I – R Code to clean Patient Survival Dataset

The following code is used to clean the Patient Survival Prediction dataset. This data cleaning process standardized for all methodologies and to ensure each team member has the same dataset to work on.

````
```{r}
# Dataset cleaning
library(dplyr)
# Factor Analysis on Patient Survival Dataset
dataset = read.csv(file="dataset.csv", header=TRUE, sep=",")
dim(dataset)
str(dataset)
#names(dataset)
#Remove ID columns
datasetv1 = dataset[,-c(1,2,3,13)]
#str(datasetv1)
datasetv2 = datasetv1[,which(colMeans(!is.na(datasetv1))>0.5)]
sum(is.na(datasetv2))
 NACheck = lapply(datasetv1, is.na)
NACheckSum = lapply(NACheck, sum)
cbind(NACheckSum)
NACheckv2 = lapply(datasetv2, is.na)
#NACheckv2
NACheckSumv2 = lapply(NACheckv2, sum)
#NACheckSumv2
cbind(NACheckSumv2)
#Listwise Deletion for variables with less than 50% of NAs in their rows.
datasetv3 <- na.omit(datasetv2)
#dim(datasetv3)
sum(is.na(datasetv3))
#str(datasetv3)
datasetv4 <- datasetv3 %>% select(where(~!is.character(.))) %>% glimpse()
#str(datasetv4)
### Remove binary columns except target column (hospital_death)
datasetv5 <- datasetv4 %>% select(-hospital_death, -elective_surgery,-readmission_status,-
apache_post_operative,-arf_apache,
                          -gcs_unable_apache,-intubated_apache,-ventilated_apache,-aids,
                          -cirrhosis,-diabetes_mellitus,-hepatic_failure,-immunosuppression,-
                       leukemia, -lymphoma,-solid_tumor_with_metastasis) %>% glimpse()
#str(datasetv5)
#dim(datasetv5)
#### Remove non continuous numerical variables
datasetv6 <- datasetv5[,-c(6,7,10,11,12)]
str(datasetv6)
dim(datasetv6)
head(datasetv6)
````
````

```

## Appendix II - Multiple Linear Regression Code

```
#Backwards selection
library(MASS)
step <- stepAIC(model, direction="backward")
step$anova #Takes a long time to run, but removes 32 variables and leaves 48.

####5-fold cross validation####
library(latticeExtra)
library(DAAG)
out <- cv.lm(data = tempData4, form.lm = formula(hospital_death ~ .),
plotit="Observed", m=5)
summary(out)

#Build Linear Model
model <- lm(tempData4$hospital_death ~ ., data = tempData4)
summary(model)

#Check for multicollinearity
library(car)
vif(model) #lots of multicollinearity exists in the model. Need to remove some
variables.

####Build linear model####

#Create new model using results of backwards selection
model2 <- lm(tempData4$hospital_death ~ age + bmi + height + pre_icu_los_days +
              weight + bun_apache + glucose_apache + hematocrit_apache +
              map_apache + resprate_apache + wbc_apache + d1_diasbp_min +
              d1_diasbp_noninvasive_min + d1_heartrate_max + d1_heartrate_min +
              d1_mbp_min + d1_resprate_min + d1_spo2_max + d1_spo2_min +
              d1_sysbp_min + d1_temp_max + d1_temp_min + h1_diasbp_max +
              h1_heartrate_max + h1_heartrate_min + h1_mbp_noninvasive_min +
              h1_resprate_min + h1_spo2_min + h1_sysbp_noninvasive_max +
              h1_temp_max + d1_bun_min + d1_calcium_max + d1_calcium_min +
              d1_creatinine_max + d1_glucose_max + d1_glucose_min + d1_hco3_max +
              d1_hco3_min + d1_hemaglobin_max + d1_hemaglobin_min + d1_hematocrit_max
+
              d1_platelets_max + d1_platelets_min + d1_sodium_max + d1_sodium_min +
              d1_wbc_max + apache_4a_hospital_death_prob + apache_4a_icu_death_prob,
            data = tempData4
)
summary(model2)
vif(model2) #multicollinearity still exists, so can remove some variables

#Remove the variables with high VIF values and create a new model.
model3 <- lm(tempData4$hospital_death ~ age + bmi + height + pre_icu_los_days +
glucose_apache +
              map_apache + resprate_apache + d1_heartrate_max + d1_heartrate_min +
              d1_mbp_min + d1_resprate_min + d1_spo2_max + d1_spo2_min +
              d1_sysbp_min + d1_temp_max + d1_temp_min + h1_diasbp_max +
```
```

```
              h1_heartrate_max + h1_heartrate_min + h1_mbp_noninvasive_min +
              h1_resprate_min + h1_spo2_min + h1_sysbp_noninvasive_max +
              h1_temp_max + d1_calcium_max + d1_calcium_min +
              d1_creatinine_max + d1_glucose_max + d1_glucose_min + d1_hco3_max +
              d1_hco3_min + d1_sodium_max + d1_sodium_min +
apache_4a_hospital_death_prob + apache_4a_icu_death_prob,
            data = tempData4
)
vif(model3) #Multicollinearity no longer exists, values look great.
summary(model3)

#Remove variables with high p-values in t-tests
model4 <- lm(tempData4$hospital_death ~ age + height + pre_icu_los_days
            + resprate_apache + d1_heartrate_max +
              d1_mbp_min + d1_resprate_min + d1_spo2_max + d1_spo2_min +
              d1_sysbp_min + d1_temp_max + d1_temp_min + h1_diasbp_max +
              h1_heartrate_max + h1_mbp_noninvasive_min +
              h1_resprate_min + h1_spo2_min + h1_temp_max +
              d1_creatinine_max + d1_hco3_max +
              d1_hco3_min + d1_sodium_max + d1_sodium_min +
apache_4a_hospital_death_prob + apache_4a_icu_death_prob,
            data = tempData4
)

vif(model4) #there is no multicollinearity
summary(model4)
plot(model4)

###See if the model can improve using Ridge and LASSO instead###

####Ridge and LASSO Regression Modelling####
x <- as.matrix(tempData4[,2:80]) #gives all rows, but subset of columns: only 2-80.
Saved as matrix
y <- as.double(tempData4[,1]) #give all rows, but only first column. Saved as double.

###Ridge Regression
library(glmnet)
set.seed(123)
ridge <- cv.glmnet (x, y, family="gaussian", alpha=0) #alpha is 0
coef(ridge, s=ridge$lambda.min) #coefficients. Keeps all the variables.
plot(ridge) #plot
ridge$lambda.min #value of Lambda

#Plotting using all the variables is back to our original model.
model_Ridge <- lm(tempData4$hospital_death ~ ., data = tempData4)
summary(model_Ridge)
plot(model_Ridge) #variables are too fitted to the model

###LASSO Regression
set.seed(123)
lasso <- cv.glmnet (x, y, family="gaussian", alpha=1) #alpha is 1
coef(lasso, s=lasso$lambda.min) #coefficients. Removes certain variables
plot(lasso) #plot
lasso$lambda.min #value of Lambda
```

```r
#Building a new model with removed variables. 18 were removed.
model_LASSO <- lm(tempData4$hospital_death ~ age + height + pre_icu_los_days +
                   weight + bun_apache + glucose_apache + hematocrit_apache +
                   map_apache + resprate_apache + temp_apache + wbc_apache +
d1_diasbp_min +
                   d1_diasbp_noninvasive_max + d1_heartrate_max + d1_heartrate_min +
                   d1_mbp_min + d1_resprate_min + d1_spo2_max + d1_spo2_min +
                   d1_sysbp_min + d1_sysbp_noninvasive_max + d1_temp_max + d1_temp_min
+ h1_diasbp_max + h1_diasbp_min
                   + h1_diasbp_noninvasive_max + h1_heartrate_max + h1_heartrate_min +
h1_mbp_max + h1_mbp_min + h1_mbp_noninvasive_min +
                   h1_resprate_max + h1_resprate_min + h1_spo2_min +
h1_sysbp_noninvasive_max +
                   h1_temp_max + d1_bun_min + d1_calcium_max + d1_calcium_min +
                   d1_creatinine_max + d1_glucose_max + d1_glucose_min + d1_hco3_max +
                   d1_hco3_min + d1_hemaglobin_max + d1_hemaglobin_min +
d1_hematocrit_max + d1_hematocrit_min +
                   d1_platelets_max + d1_platelets_min + d1_potassium_max +
d1_potassium_min + d1_sodium_max + d1_sodium_min +
                   d1_wbc_max + apache_4a_hospital_death_prob +
apache_4a_icu_death_prob,
                 data = tempData4
)
summary(model_LASSO)
plot(model_LASSO)

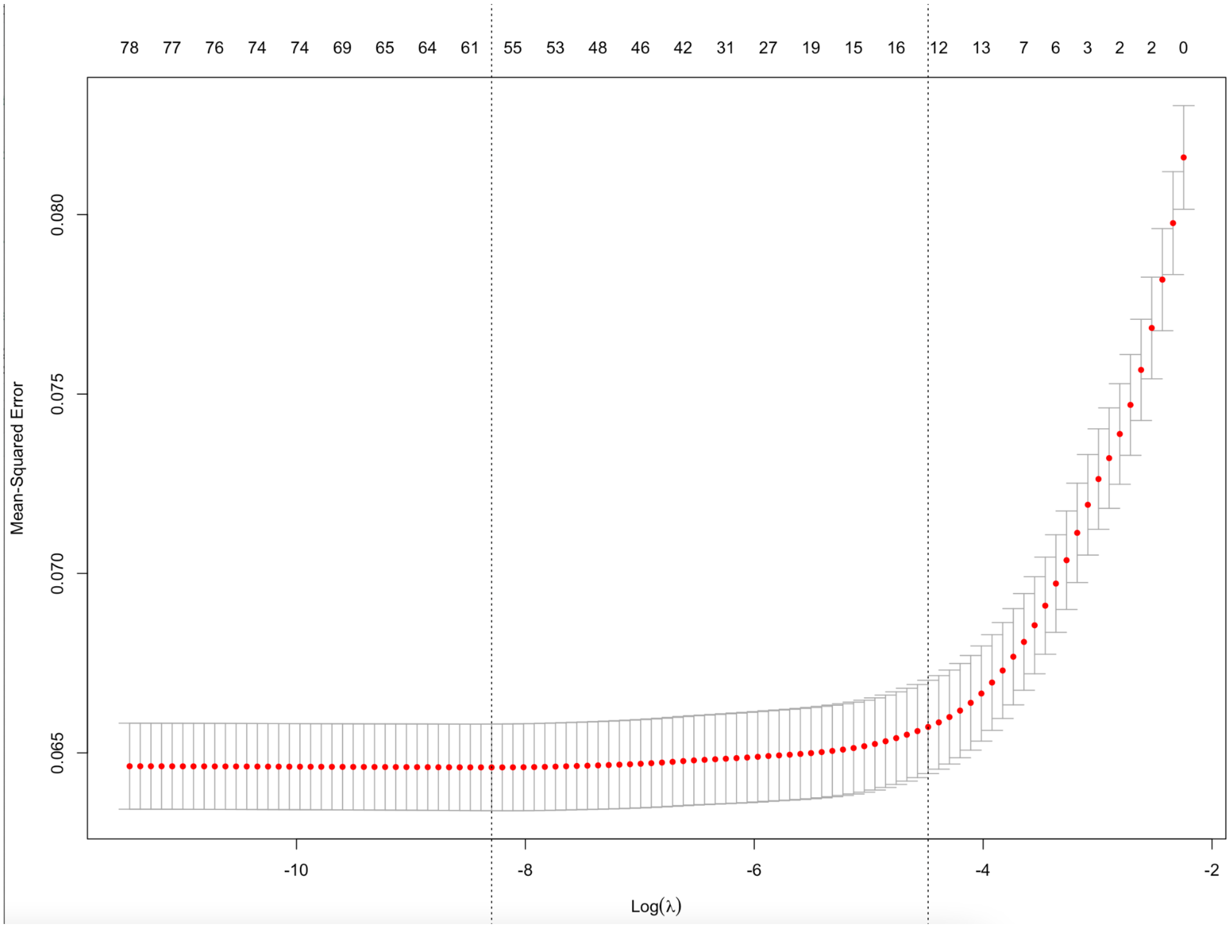
#Histogram to see distribution of ages
ages <- tempData4$age
hist(ages)

#Frequency table of deaths
factor(tempData4$hospital_death)
table(tempData4$hospital_death) #31,657 alive, 3119 deaths

#dplyr frequency table
library(plyr)
count(tempData4, "age")
count(tempData4, "age", "hospital_death")

#Cross Tabulation using gmodels
library(gmodels)
CrossTable(tempData4$hospital_death, prop.t=TRUE, prop.r=TRUE, prop.c=TRUE)
```

Appendix II.I - Multiple Linear Regression Outputs



*LASSO Regression Plot (*Figure 1)*

Appendix III – Factor Analysis Code

````
```{r }
# Factorability / Reliability Tests
#Test KMO Sampling Adequacy
library(psych)
KMO(datasetv6)
#Overall MSA =  0.81

#Test Bartlett's Test of Sphericity => checking for correlation
library(REdaS)
bart_spher(datasetv6)
#p-value < 2.22e-16 (Very Small Number)

#Test for Reliability Analysis using Cronbach's Alpha => Looking for >0.7 => these data points
belong together
alpha(datasetv6,check.keys=TRUE)
#raw_alpha = 0.8199

# Create a Scree plot to see the number of factors to use
p = prcomp(datasetv6, center=T, scale=T)
p
#Check Scree Plot
plot(p)
abline(1, 0)
summary(p)
#options(max.print = 10000)
#print(p)
table(p$sdev>1)

library(psych)
#Conducting Factor Analysis
fit = psych::fa(datasetv6, rotate="varimax", nfactors=5, scores=TRUE)
print(fit$loadings, cutoff=.4, sort=T)
summary(fit)

library(dplyr)
library(ggplot2)
# Factor loadings separation
fa_factors <- as.data.frame(unclass(fit$loadings))
fa_factor1_loading <- as.data.frame(fit$loadings[,1])
````
````

```
fa_factor1_loading_big_point4 <- fa_factor1_loading %>% filter_at(vars(1:1), any_vars(abs(.)
>0.4))
fa_factor5_loading <- as.data.frame(fit$loadings[,2])
fa_factor5_loading_big_point4 <- fa_factor5_loading %>% filter_at(vars(1:1), any_vars(abs(.)
>0.4))
fa_factor2_loading <- as.data.frame(fit$loadings[,3])
fa_factor2_loading_big_point4 <- fa_factor2_loading %>% filter_at(vars(1:1), any_vars(abs(.)
>0.4))
fa_factor4_loading <- as.data.frame(fit$loadings[,4])
fa_factor4_loading_big_point4 <- fa_factor4_loading %>% filter_at(vars(1:1), any_vars(abs(.)
>0.4))
fa_factor3_loading <- as.data.frame(fit$loadings[,5])
fa_factor3_loading_big_point4 <- fa_factor3_loading %>% filter_at(vars(1:1), any_vars(abs(.)
>0.4))

#horizontal bar plot Factor 1
ggplot(fa_factor1_loading_big_point4, aes(x=rownames(fa_factor1_loading_big_point4),
y=fa_factor1_loading_big_point4[,1]))+
  geom_bar(stat = "identity") + coord_flip() + ggtitle("Factor 1 loadings bigger than 0.4") +
  xlab("Loadings") + ylab("Variables")

#horizontal bar plot Factor 2
ggplot(fa_factor2_loading_big_point4, aes(x=rownames(fa_factor2_loading_big_point4),
y=fa_factor2_loading_big_point4[,1]))+
  geom_bar(stat = "identity") + coord_flip() + ggtitle("Factor 2 loadings bigger than 0.4") +
  xlab("Loadings") + ylab("Variables")

#horizontal bar plot Factor 3
ggplot(fa_factor3_loading_big_point4, aes(x=rownames(fa_factor3_loading_big_point4),
y=fa_factor3_loading_big_point4[,1]))+
  geom_bar(stat = "identity") + coord_flip() + ggtitle("Factor 3 loadings bigger than 0.4") +
  xlab("Loadings") + ylab("Variables")
#horizontal bar plot Factor 4

ggplot(fa_factor4_loading_big_point4, aes(x=rownames(fa_factor4_loading_big_point4),
y=fa_factor4_loading_big_point4[,1]))+
  geom_bar(stat = "identity") + coord_flip() + ggtitle("Factor 4 loadings bigger than 0.4") +
  xlab("Loadings") + ylab("Variables")

#horizontal bar plot Factor 5
```

```
ggplot(fa_factor5_loading_big_point4, aes(x=rownames(fa_factor5_loading_big_point4), 
y=fa_factor5_loading_big_point4[,1]))+
  geom_bar(stat = "identity") + coord_flip() + ggtitle("Factor 5 loadings bigger than 0.4") +
  xlab("Loadings") + ylab("Variables")
```

Appendix IV - Factor Loadings Figure

Loadings:

	MR1	MR5	MR2	MR4	MR3
map_apache	0.533				
d1_diasbp_max	0.767				
d1_diasbp_noninvasive_max	0.768				
d1_mbp_max	0.858				
d1_mbp_noninvasive_max	0.856				
d1_sysbp_max	0.791				
d1_sysbp_noninvasive_max	0.792				
h1_diasbp_max	0.772				
h1_diasbp_noninvasive_max	0.769				
h1_mbp_max	0.846				
h1_mbp_min	0.612	0.606			
h1_mbp_noninvasive_max	0.847				
h1_mbp_noninvasive_min	0.612	0.607			
h1_sysbp_max	0.790				
h1_sysbp_noninvasive_max	0.793				
d1_diasbp_min		0.721			
d1_diasbp_noninvasive_min		0.721			
d1_mbp_min		0.785			
d1_mbp_noninvasive_min		0.786			
d1_sysbp_min		0.756			
d1_sysbp_noninvasive_min		0.757			
h1_diasbp_min	0.520	0.556			
h1_diasbp_noninvasive_min	0.521	0.558			
h1_sysbp_min	0.573	0.578			
h1_sysbp_noninvasive_min	0.575	0.578			
bun_apache			0.864		
creatinine_apache			0.849		
d1_bun_max			0.864		
d1_bun_min			0.838		
d1_creatinine_max			0.848		
d1_creatinine_min			0.833		
hematocrit_apache				0.917	
d1_hemaglobin_max				0.874	
d1_hemaglobin_min				0.900	
d1_hematocrit_max				0.909	
d1_hematocrit_min				0.916	
heart_rate_apache					0.724
d1_heartrate_max					0.716
d1_heartrate_min					0.633
h1_heartrate_max					0.787
h1_heartrate_min					0.802
age					
bmi					
height					
pre_icu_los_days					
weight					
glucose_apache					
resprate_apache					
sodium_apache					
temp_apache					
wbc_apache					
d1_resprate_max					
d1_resprate_min					
d1_spo2_max					
d1_spo2_min					
d1_temp_max					
d1_temp_min					
h1_resprate_max					
h1_resprate_min					0.408
h1_spo2_max					
h1_spo2_min					
h1_temp_max					0.424
h1_temp_min					0.415
d1_calcium_max					
d1_calcium_min					
d1_glucose_max					
d1_glucose_min					
d1_hco3_max					
d1_hco3_min					
d1_platelets_max					
d1_platelets_min					
d1_potassium_max			0.429		
d1_potassium_min					
d1_sodium_max					
d1_sodium_min					
d1_wbc_max					
d1_wbc_min					

Appendix V - Principal Component Analysis Code

```
library(DescTools)
library(psych)
library(REdaS)
finalData <- datasetv6
# Removing problematic variables, high kurtosis, low communality, negatively skewed
finalData <- datasetv6 %>% select(-h1_temp_max,-h1_temp_min,-h1_spo2_max,
                          -h1_spo2_min,-h1_resprate_min,
                          -h1_resprate_max,-d1_temp_min,-d1_temp_max,
                          -d1_spo2_max,-d1_spo2_min,-temp_apache) %>% glimpse()

cols <- c("bmi","pre_icu_los_days","bun_apache","creatinine_apache",
      "glucose_apache","wbc_apache","d1_resprate_max",
      "d1_bun_max","d1_bun_min","d1_creatinine_max","d1_creatinine_min",
      "d1_glucose_max","d1_wbc_max","d1_wbc_min",
      "apache_4a_hospital_death_prob","apache_4a_icu_death_prob"
)
finalData[cols] <- log(finalData[cols])

describe(finalData)

finalData <- na.omit(finalData)
sum(is.na(finalData))

finalData <- finalData[!is.infinite(rowSums(finalData)),]

KMO(finalData)
#Overall MSA =  0.82

#Test Bartlett's Test of Sphericity
bart_spher(finalData)
#p-value < 2.22e-16 (Very Small Number)

alpha(finalData,check.keys=TRUE)
# Raw alpha = 0.86

describe(finalData)

library(factoextra)
library(FactoMineR)
```

```
p = prcomp(finalData, center=T, scale=T)
p

summary(p)

p2 <- prcomp(finalData, scale = TRUE)
fviz_eig(p2)

p3 = psych::principal(finalData, rotate="varimax", nfactors=4, scores=TRUE)
p3
print(p3$loadings, cutoff=.422, sort=T)

p3$communality
p3$rot.mat

table(p3$values>1)

finalData2 <- finalData %>% select(-d1_sodium_max,-d1_sodium_min,-d1_potassium_min,-
d1_potassium_max,
                         -d1_platelets_max,-d1_platelets_min,-d1_hco3_max,
                         -d1_hco3_min,-d1_glucose_max,-d1_glucose_min,
                         -d1_calcium_max,-d1_calcium_min,-d1_resprate_max,
                         -d1_resprate_min,-sodium_apache,-resprate_apache,-glucose_apache,
                         -weight,-pre_icu_los_days,-height,-bmi,-age,-wbc_apache,-d1_wbc_max,-
d1_wbc_min) %>% glimpse()

p4 <- prcomp(finalData2, scale = TRUE)
fviz_eig(p4, linecolor="red",barcolor="darkblue",barfill="darkblue")

p5 = psych::principal(finalData2, rotate="varimax", nfactors=4, scores=TRUE)
p5
print(p5$loadings, cutoff=.422, sort=T)
ls(p5)

p6 <- PCA(finalData2, ncp = 4, graph = FALSE)

variables <- get_pca_var(p4)

# Contributions of variables to PC1
```

```
fviz_contrib(p4, choice = "var", axes = 1, top = 10, color="darkblue",fill="darkblue") +
labs(title='Contributions of variables to PC1')
# Contributions of variables to PC2
fviz_contrib(p4, choice = "var", axes = 2, top = 10, color="darkblue",fill="darkblue") +
labs(title='Contributions of variables to PC2')
# Contributions of variables to PC3
fviz_contrib(p4, choice = "var", axes = 3, top = 10, color="darkblue",fill="darkblue") +
labs(title='Contributions of variables to PC3')
# Contributions of variables to PC4
fviz_contrib(p4, choice = "var", axes = 4, top = 10, color="darkblue",fill="darkblue") +
labs(title='Contributions of variables to PC4')

scores <- p3$scores
summary(scores)
```

Appendix VI - PCA Component Loadings

	RC1	RC2	RC4	RC3
map_apache	0.573			
d1_diasbp_max	0.583			
d1_diasbp_min	0.608			
d1_diasbp_noninvasive_max	0.583			
d1_diasbp_noninvasive_min	0.608			
d1_mbp_max	0.697			
d1_mbp_min	0.680			
d1_mbp_noninvasive_max	0.697			
d1_mbp_noninvasive_min	0.680			
d1_sysbp_max	0.705			
d1_sysbp_min	0.654			
d1_sysbp_noninvasive_max	0.705			
d1_sysbp_noninvasive_min	0.654			
h1_diasbp_max	0.729			
h1_diasbp_min	0.756			
h1_diasbp_noninvasive_max	0.727			
h1_diasbp_noninvasive_min	0.757			
h1_mbp_max	0.833			
h1_mbp_min	0.847			
h1_mbp_noninvasive_max	0.834			
h1_mbp_noninvasive_min	0.848			
h1_sysbp_max	0.801			
h1_sysbp_min	0.793			
heart_rate_apache				0.748
d1_heartrate_max				0.804
d1_heartrate_min				0.506
h1_heartrate_max				0.786
h1_heartrate_min				0.699
apache_4a_icu_death_prob		0.498		
d1_creatinine_max		0.831		
d1_creatinine_min		0.819		
apache_4a_hospital_death_prob		0.513		
hematocrit_apache			0.819	
d1_hemaglobin_max			0.809	
d1_hemaglobin_min			0.809	
d1_hematocrit_max			0.832	
d1_hematocrit_min			0.818	

AppendixVII- Canonical Correlation Analysis Code and Figures

```
##CCA model using yacca package

ccaModel = cca(demoVars,vitalVars)
summary(ccaModel)

#normalize data demo
hospital_death <- tempData3$hospital_death

normAge <- as.data.frame(tempData3$age)
MinMaxAge <- preProcess(normAge, method=c("range"))
newges <- predict(MinMaxAge, normAge)
summary(newAges)

normbmi <- as.data.frame(tempData3$bmi)
MinMaxbmi <- preProcess(normbmi, method=c("range"))
newbmi <- predict(MinMaxbmi, normbmi)
summary(newbmi)

normheight <- as.data.frame(tempData3$height)
MinMaxheight <- preProcess(normheight, method=c("range"))
newheight <- predict(MinMaxheight, normheight)
summary(newheight)

normpre_icu_los_days <- as.data.frame(tempData3$pre_icu_los_days)
MinMaxpre_icu_los_days <- preProcess(normpre_icu_los_days, method=c("range"))
newpre_icu_los_days <- predict(MinMaxpre_icu_los_days, normpre_icu_los_days)
summary(newpre_icu_los_days)

normweight <- as.data.frame(tempData3$weight)
MinMaxweight <- preProcess(normweight, method=c("range"))
newweight <- predict(MinMaxweight, normweight)
summary(newweight)

#normalized variable sets
demoVarsNorm <- data.frame(newAges,newbmi,
newheight,newpre_icu_los_days,newweight, hospital_death)
summary(demoVarsNorm)

#normalize data vitals
```

normVitalVars <- as.data.frame(vitalVars)
MinMaxVitalVars <- preProcess(normVitalVars, method=c("range"))
vitalVarsNorm <- predict(MinMaxVitalVars, normVitalVars)

summary(vitalVarsNorm)

#CCA
ccaModel2 = cca(demoVarsNorm,vitalVarsNorm)
summary(ccaModel2)

AppendixVIII - CCA and Summary Statistic R Output

```
> demoVars <- tempData3 %>% select(1:6)
> summary(demoVars)
 hospital_death         age             bmi            height       pre_icu_los_days      weight
 Min.   :0.00000   Min.   :16.00   Min.   :14.84   Min.   :137.2   Min.   :-0.2243   Min.   : 38.60
 1st Qu.:0.00000   1st Qu.:53.00   1st Qu.:23.66   1st Qu.:162.6   1st Qu.: 0.0250   1st Qu.: 67.10
 Median :0.00000   Median :65.00   Median :27.71   Median :170.2   Median : 0.1278   Median : 80.80
 Mean   :0.08969   Mean   :62.58   Mean   :29.22   Mean   :169.9   Mean   : 0.7846   Mean   : 84.41
 3rd Qu.:0.00000   3rd Qu.:75.00   3rd Qu.:33.02   3rd Qu.:177.8   3rd Qu.: 0.3722   3rd Qu.: 97.70
 Max.   :1.00000   Max.   :89.00   Max.   :67.81   Max.   :195.6   Max.   :67.0236   Max.   :186.00
> #separating vitals variables for analysis and looking at summary for normality
> vitalVars <- tempData3 %>% select(22:41)
> summary(vitalVars)
 d1_diasbp_max    d1_diasbp_min    d1_diasbp_noninvasive_max d1_diasbp_noninvasive_min d1_heartrate_max d1_heartrate_min
 Min.   : 46.0   Min.   :13.00   Min.   : 46.0             Min.   :13.00             Min.   : 58.0    Min.   :  0.00
 1st Qu.: 77.0   1st Qu.:40.00   1st Qu.: 77.0             1st Qu.:40.00             1st Qu.: 90.0    1st Qu.: 60.00
 Median : 88.0   Median :48.00   Median : 88.0             Median :48.00             Median :103.0    Median : 69.00
 Mean   : 90.6   Mean   :48.26   Mean   : 90.6             Mean   :48.26             Mean   :105.4    Mean   : 70.25
 3rd Qu.:100.0   3rd Qu.:56.00   3rd Qu.:100.0             3rd Qu.:56.00             3rd Qu.:119.0    3rd Qu.: 80.00
 Max.   :165.0   Max.   :90.00   Max.   :165.0             Max.   :90.00             Max.   :177.0    Max.   :143.00
   d1_mbp_max      d1_mbp_min      d1_mbp_noninvasive_max d1_mbp_noninvasive_min d1_resprate_max d1_resprate_min
 Min.   : 60.0   Min.   : 22.00   Min.   : 60.0          Min.   : 22.00         Min.   :14.00   Min.   : 0.00
 1st Qu.: 91.0   1st Qu.: 53.00   1st Qu.: 91.0          1st Qu.: 53.00         1st Qu.:23.00   1st Qu.:10.00
 Median :103.0   Median : 62.00   Median :103.0          Median : 62.00         Median :27.00   Median :12.00
 Mean   :105.6   Mean   : 62.37   Mean   :105.6          Mean   : 62.34         Mean   :29.46   Mean   :12.38
 3rd Qu.:117.0   3rd Qu.: 71.00   3rd Qu.:117.0          3rd Qu.: 71.00         3rd Qu.:33.00   3rd Qu.:15.00
 Max.   :184.0   Max.   :112.00   Max.   :181.0          Max.   :112.00         Max.   :92.00   Max.   :72.00
  d1_spo2_max      d1_spo2_min       d1_sysbp_max  d1_sysbp_min d1_sysbp_noninvasive_max d1_sysbp_noninvasive_min
 Min.   : 13.00   Min.   :  0.00   Min.   : 90   Min.   : 41   Min.   : 90              Min.   : 41.03
 1st Qu.: 99.00   1st Qu.: 89.00   1st Qu.:132   1st Qu.: 82   1st Qu.:132              1st Qu.: 82.00
 Median :100.00   Median : 92.00   Median :147   Median : 93   Median :147              Median : 93.00
 Mean   : 99.43   Mean   : 90.03   Mean   :150   Mean   : 94   Mean   :150              Mean   : 94.00
 3rd Qu.:100.00   3rd Qu.: 95.00   3rd Qu.:165   3rd Qu.:106   3rd Qu.:165              3rd Qu.:106.00
 Max.   :100.00   Max.   :100.00   Max.   :232   Max.   :160   Max.   :232              Max.   :160.00
  d1_temp_max      d1_temp_min
 Min.   :35.10   Min.   :31.89
 1st Qu.:36.90   1st Qu.:36.10
 Median :37.20   Median :36.40
 Mean   :37.37   Mean   :36.22
 3rd Qu.:37.70   3rd Qu.:36.60
 Max.   :39.90   Max.   :37.80
```

```
> summary(ccaModel)

Canonical Correlation Analysis - Summary


Canonical Correlations:

      CV 1       CV 2       CV 3       CV 4       CV 5       CV 6 
0.39859254 0.33228311 0.15874219 0.13218139 0.06167876 0.01795069 

Shared Variance on Each Canonical Variate:

        CV 1         CV 2         CV 3         CV 4         CV 5         CV 6 
0.1588760148 0.1104120659 0.0251990840 0.0174719206 0.0038042695 0.0003222271 

Bartlett's Chi-Squared Test:

          rho^2      Chisq  df    Pr(>X)    
CV 1 1.5888e-01 1.1725e+04 120 < 2.2e-16 ***
CV 2 1.1041e-01 5.7106e+03  95 < 2.2e-16 ***
CV 3 2.5199e-02 1.6436e+03  72 < 2.2e-16 ***
CV 4 1.7472e-02 7.5642e+02  51 < 2.2e-16 ***
CV 5 3.8043e-03 1.4370e+02  32 4.441e-16 ***
CV 6 3.2223e-04 1.1203e+01  15    0.7381    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Canonical Variate Coefficients:

        X Vars:
                          CV 1         CV 2          CV 3        CV 4         CV 5          CV 6
hospital_death   -1.8076386945 -2.889886739 -0.8378796350  0.04808231  0.293355190 -0.0523541652
age              -0.0449815001  0.034580548 -0.0007031255  0.02057321 -0.011454347 -0.0007133837
bmi              -0.0006451274  0.002142573  0.0722344674  0.03162117  0.020770122 -0.6247760984
height            0.0293949589 -0.008288487 -0.0136351887  0.08766075 -0.036073418 -0.2271897454
pre_icu_los_days  0.0074830265 -0.061814495  0.1593965469 -0.13590499 -0.372929737  0.0013727150
weight           -0.0092354750  0.011086648 -0.0510328719 -0.04243375 -0.008945876  0.2143455041

        Y Vars:
                                  CV 1          CV 2          CV 3          CV 4         CV 5        CV 6
d1_diasbp_max            -0.0043045774 -9.809180e-03 -0.0695184568  0.0458003062  0.340068904 -0.03689879
d1_diasbp_min             0.0495710620  7.231451e-03  0.0004166168 -0.1685552214  0.148254827  0.05552870
d1_diasbp_noninvasive_max 0.0203667792  8.107866e-05  0.0517042965 -0.0511478162 -0.320381051  0.08431055
d1_diasbp_noninvasive_min 0.0035490506 -3.731426e-02 -0.0429880036  0.2159039176 -0.188757466 -0.02244143
d1_heartrate_max          0.0079528337 -2.475654e-02  0.0102896613 -0.0002013993  0.005484678  0.01663073
d1_heartrate_min          0.0014946422  4.817957e-05  0.0023763106 -0.0434576728 -0.004279431 -0.01687191
d1_mbp_max               -0.0052984294 -2.367038e-03  0.0208927170 -0.0092988988  0.015123714 -0.13264213
d1_mbp_min                0.0083275196  1.279844e-02  0.0040227262 -0.0093656254  0.024181374 -0.06277013
d1_mbp_noninvasive_max   -0.0006439011  7.497766e-03 -0.0205032197  0.0025537983  0.014804126  0.06177692
d1_mbp_noninvasive_min   -0.0077608811 -4.059428e-03  0.0086323225  0.0147128596 -0.002609622  0.03022660
d1_resprate_max          -0.0086329276 -5.070727e-03 -0.0030679185 -0.0152629468 -0.013501696  0.03637746
d1_resprate_min          -0.0517449173 -8.514981e-03  0.0428115671  0.0203375606  0.083671911  0.04403343
d1_spo2_max               0.1023655478 -9.051283e-02  0.4633890949  0.1369757980  0.066657816  0.05358280
d1_spo2_min               0.0300070104  2.221042e-02  0.0319627533  0.0015409852  0.029821486  0.02206003
d1_sysbp_max             -0.0045800089  6.517405e-03  0.0010728506 -0.0477999343 -0.302312535  0.14105358
d1_sysbp_min              0.0044568149  9.701694e-03 -0.0369248335  0.0748090768 -0.280842590  0.06900246
d1_sysbp_noninvasive_max -0.0147224317  6.073962e-03 -0.0083691822  0.0488923911  0.280438982 -0.10741431
d1_sysbp_noninvasive_min  0.0012999358  6.202005e-04  0.0374979930 -0.1038703343  0.290700345 -0.05993974
d1_temp_max               0.1773291438 -2.782451e-01 -0.4062190749 -0.3198619018 -0.290193323 -0.30897867
d1_temp_min               0.3368154733  6.615332e-01  0.3124872066 -0.2493440461 -0.280990709  0.16711361
```

```
Structural Correlations (Loadings):

        X Vars:
                        CV 1        CV 2       CV 3        CV 4        CV 5         CV 6
hospital_death   -0.585438404 -0.78508744 -0.1955008  0.04282216  0.01913764 -0.021887424
age              -0.801330932  0.45763583  0.1099088  0.32936142 -0.16488790  0.026292190
bmi              -0.152196715  0.23843355 -0.5152485 -0.75495520  0.01125402 -0.290552163
height            0.316750371 -0.03836157 -0.6773968  0.47984313 -0.45193947 -0.069515418
pre_icu_los_days -0.051909742 -0.16637180  0.3665945 -0.30169623 -0.86267570 -0.002080888
weight           -0.005315431  0.21050840 -0.8040515 -0.50571715 -0.19482607 -0.124306460

        Y Vars:
                                 CV 1        CV 2        CV 3        CV 4        CV 5        CV 6
d1_diasbp_max               0.05371354 -0.04056083 -0.48527819 -0.16465907  0.56591547  0.28001875
d1_diasbp_min               0.72335238  0.06645865 -0.41448637  0.19227062  0.04302282  0.08995916
d1_diasbp_noninvasive_max   0.05384501 -0.03967147 -0.48363846 -0.16342277  0.56012972  0.28327448
d1_diasbp_noninvasive_min   0.72240217  0.06554580 -0.41218325  0.19450968  0.03889720  0.08872038
d1_heartrate_max            0.09464275 -0.69607193  0.11572398 -0.40941739  0.04597911  0.23743869
d1_heartrate_min            0.24554532 -0.30137382  0.13978761 -0.73163371 -0.02414440 -0.05970063
d1_mbp_max                 -0.04333594  0.09300864 -0.48568714 -0.16332278  0.51356117  0.02455702
d1_mbp_min                  0.57055745  0.25825047 -0.33913532  0.06884529  0.14566637  0.01877894
d1_mbp_noninvasive_max     -0.04135247  0.09917182 -0.49566556 -0.15757258  0.51827913  0.07959109
d1_mbp_noninvasive_min      0.57124807  0.25634211 -0.33666951  0.06918973  0.14353838  0.01689414
d1_resprate_max            -0.18818245 -0.22389324 -0.05456081 -0.23546435 -0.08601822  0.28027772
d1_resprate_min            -0.12980866 -0.03192264  0.16400246 -0.12585045  0.32853979  0.22261297
d1_spo2_max                 0.10128400 -0.17498866  0.60048409  0.17101377  0.06357975  0.09029787
d1_spo2_min                 0.43668693  0.36746273  0.32152755  0.05023964  0.25637814  0.13682294
d1_sysbp_max               -0.23638571  0.30983863 -0.42539617 -0.14404022  0.17529145  0.37891750
d1_sysbp_min                0.40076255  0.45227659 -0.26081563 -0.13858899  0.16136221  0.13348701
d1_sysbp_noninvasive_max   -0.23596856  0.31125246 -0.42590858 -0.14193574  0.18885082  0.37716566
d1_sysbp_noninvasive_min    0.40006701  0.45080814 -0.25805589 -0.13898624  0.16791484  0.13051365
d1_temp_max                 0.15655022 -0.24583312 -0.08908824 -0.42547583 -0.27155677 -0.09488762
d1_temp_min                 0.38096710  0.48045422  0.16360450 -0.37325622 -0.21335821  0.08250143


Fractional Variance Deposition on Canonical Variates:

        X Vars:
                         CV 1       CV 2       CV 3        CV 4         CV 5         CV 6
hospital_death   3.427381e-01 0.61636228 0.03822054 0.001833738 0.0003662494 4.790593e-04
age              6.421313e-01 0.20943056 0.01207993 0.108478948 0.0271880207 6.912793e-04
bmi              2.316384e-02 0.05685056 0.26548103 0.569957357 0.0001266530 8.442056e-02
height           1.003308e-01 0.00147161 0.45886648 0.230249432 0.2042492816 4.832393e-03
pre_icu_los_days 2.694621e-03 0.02767958 0.13439149 0.091020617 0.7442093635 4.330096e-06
weight           2.825381e-05 0.04431379 0.64649883 0.255749831 0.0379571964 1.545210e-02

        Y Vars:
                                  CV 1        CV 2        CV 3        CV 4         CV 5         CV 6
d1_diasbp_max              0.002885144 0.001645181 0.235494924 0.027112609 0.3202603197 0.0784104990
d1_diasbp_min              0.523238668 0.004416753 0.171798955 0.036967990 0.0018509627 0.0080926499
d1_diasbp_noninvasive_max  0.002899285 0.001573825 0.233906160 0.026707001 0.3137453031 0.0802444284
d1_diasbp_noninvasive_min  0.521864890 0.004296252 0.169895033 0.037834016 0.0015129918 0.0078713066
d1_heartrate_max           0.008957251 0.484516125 0.013392040 0.167622602 0.0021140783 0.0563771305
d1_heartrate_min           0.060292502 0.090826179 0.019540576 0.535287881 0.0005829521 0.0035641658
d1_mbp_max                 0.001878004 0.008650608 0.235892000 0.026674331 0.2637450804 0.0006030470
d1_mbp_min                 0.325535800 0.066693308 0.115012764 0.004739675 0.0212186922 0.0003526487
d1_mbp_noninvasive_max     0.001710027 0.009835050 0.245684344 0.024829118 0.2686132591 0.0063347410
d1_mbp_noninvasive_min     0.326324353 0.065711275 0.113346361 0.004787218 0.0206032669 0.0002854119
d1_resprate_max            0.035412636 0.050128182 0.002976882 0.055443462 0.0073991350 0.0785556017
d1_resprate_min            0.016850287 0.001019055 0.026896807 0.015838337 0.1079383960 0.0495565339
d1_spo2_max                0.010258449 0.030621031 0.360581137 0.029245710 0.0040423843 0.0081537058
d1_spo2_min                0.190695478 0.135028856 0.103379964 0.002524022 0.0657297527 0.0187205158
d1_sysbp_max               0.055878202 0.095999974 0.180961901 0.020747586 0.0307270908 0.1435784725
d1_sysbp_min               0.160610619 0.204554112 0.068024792 0.019206908 0.0260377627 0.0178187810
d1_sysbp_noninvasive_max   0.055681159 0.096878095 0.181398121 0.020145753 0.0356646309 0.1422539354
d1_sysbp_noninvasive_min   0.160053615 0.203227979 0.066592842 0.019317176 0.0281953921 0.0170338127
d1_temp_max                0.024507971 0.060433925 0.007936714 0.181029684 0.0737430786 0.0090036595
d1_temp_min                0.145135928 0.230836255 0.026766433 0.139320209 0.0455217279 0.0068064855
```

```
Canonical Communalities (Fraction of Total Variance
Explained for Each Variable, Within Sets):

        X Vars:
  hospital_death              age              bmi           height pre_icu_los_days           weight
               1                1                1                1                1                1

        Y Vars:
            d1_diasbp_max             d1_diasbp_min d1_diasbp_noninvasive_max d1_diasbp_noninvasive_min
                0.6658087                 0.7463660                 0.6590760                 0.7432745
         d1_heartrate_max          d1_heartrate_min                d1_mbp_max                d1_mbp_min
                0.7329792                 0.7100943                 0.5374431                 0.5335529
   d1_mbp_noninvasive_max    d1_mbp_noninvasive_min           d1_resprate_max           d1_resprate_min
                0.5570065                 0.5310579                 0.2299159                 0.2180994
              d1_spo2_max               d1_spo2_min             d1_sysbp_max              d1_sysbp_min
                0.4429024                 0.5160786                 0.5278932                 0.4962530
 d1_sysbp_noninvasive_max  d1_sysbp_noninvasive_min               d1_temp_max               d1_temp_min
                0.5320217                 0.4944208                 0.3566550                 0.5943870

Canonical Variate Adequacies (Fraction of Total Variance
Explained by Each CV, Within Sets):


        X Vars:
      CV 1       CV 2       CV 3       CV 4       CV 5       CV 6
0.18518115 0.15935140 0.25925639 0.20954832 0.16901613 0.01764662

        Y Vars:
      CV 1       CV 2       CV 3       CV 4       CV 5       CV 6
0.13153351 0.09234460 0.12897394 0.06976906 0.08196231 0.03668088

Redundancy Coefficients (Fraction of Total Variance
Explained by Each CV, Across Sets):


        X | Y:
        CV 1         CV 2         CV 3         CV 4         CV 5         CV 6
2.942084e-02 1.759432e-02 6.533023e-03 3.661212e-03 6.429829e-04 5.686220e-06

        Y | X:
        CV 1         CV 2         CV 3         CV 4         CV 5         CV 6
2.089752e-02 1.019596e-02 3.250025e-03 1.219000e-03 3.118067e-04 1.181957e-05

Aggregate Redundancy Coefficients (Total Variance
Explained by All CVs, Across Sets):

        X | Y: 0.05785806
        Y | X: 0.03588613
```