HW2

Caitlin Bolz

2/25/2021

library(dplyr)

library(fastDummies)  
library(mice)

# Part 1

rhc = read.table("rhcdata.txt", header = T, row.names = 1)

Looking at the structure of the data and which variables have missing values.

str(rhc)

## 'data.frame': 5735 obs. of 63 variables:  
## $ cat1 : chr "COPD" "MOSF w/Sepsis" "MOSF w/Malignancy" "ARF" ...  
## $ cat2 : chr NA NA "MOSF w/Sepsis" NA ...  
## $ ca : chr "Yes" "No" "Yes" "No" ...  
## $ sadmdte : int 11142 11799 12083 11146 12035 12389 12381 11453 12426 11381 ...  
## $ dschdte : int 11151 11844 12143 11183 12037 12396 12423 11487 12437 11400 ...  
## $ dthdte : int NA 11844 NA 11183 12037 NA NA 11491 NA NA ...  
## $ lstctdte: int 11382 11844 12400 11182 12036 12590 12616 11490 12560 11590 ...  
## $ death : int 0 1 0 1 1 0 0 1 0 0 ...  
## $ cardiohx: int 0 1 0 0 0 0 0 0 0 0 ...  
## $ chfhx : int 0 1 0 0 0 1 0 0 0 0 ...  
## $ dementhx: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ psychhx : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ chrpulhx: int 1 0 0 0 0 1 0 0 0 0 ...  
## $ renalhx : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ liverhx : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ gibledhx: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ malighx : int 1 0 1 0 0 0 1 0 0 1 ...  
## $ immunhx : int 0 1 1 1 0 0 0 0 0 0 ...  
## $ transhx : int 0 1 0 0 0 0 0 1 0 0 ...  
## $ amihx : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ age : num 70.3 78.2 46.1 75.3 67.9 ...  
## $ sex : chr "Male" "Female" "Female" "Female" ...  
## $ edu : num 12 12 14.07 9 9.95 ...  
## $ surv2md1: num 0.641 0.755 0.317 0.441 0.437 ...  
## $ das2d3pc: num 23.5 14.8 18.1 22.9 21.1 ...  
## $ t3d30 : int 30 30 30 30 2 30 30 30 30 30 ...  
## $ dth30 : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ aps1 : int 46 50 82 48 72 38 29 25 47 48 ...  
## $ scoma1 : int 0 0 0 0 41 0 26 100 0 0 ...  
## $ meanbp1 : num 41 63 57 55 65 115 67 128 53 73 ...  
## $ wblc1 : num 22.1 28.9 0.05 23.3 29.7 ...  
## $ hrt1 : int 124 137 130 58 125 134 135 102 118 141 ...  
## $ resp1 : num 10 38 40 26 27 36 10 34 30 40 ...  
## $ temp1 : num 38.7 38.9 36.4 35.8 34.8 ...  
## $ pafi1 : num 68 218 276 157 478 ...  
## $ alb1 : num 3.5 2.6 3.5 3.5 3.5 ...  
## $ hema1 : num 58 32.5 21.1 26.3 24 ...  
## $ bili1 : num 1.01 0.7 1.01 0.4 1.01 ...  
## $ crea1 : num 1.2 0.6 2.6 1.7 3.6 ...  
## $ sod1 : int 145 137 146 117 126 138 136 136 136 146 ...  
## $ pot1 : num 4 3.3 2.9 5.8 5.8 ...  
## $ paco21 : num 40 34 16 30 17 68 45 26 40 30 ...  
## $ ph1 : num 7.36 7.33 7.36 7.46 7.23 ...  
## $ swang1 : chr "NoRHC" "RHC" "RHC" "NoRHC" ...  
## $ wtkilo1 : num 64.7 45.7 NA 54.6 78.4 ...  
## $ dnr1 : chr "No" "No" "No" "No" ...  
## $ ninsclas: chr "Medicare" "Private & Medicare" "Private" "Private & Medicare" ...  
## $ resp : chr "Yes" "No" "No" "Yes" ...  
## $ card : chr "Yes" "No" "Yes" "No" ...  
## $ neuro : chr "No" "No" "No" "No" ...  
## $ gastr : chr "No" "No" "No" "No" ...  
## $ renal : chr "No" "No" "No" "No" ...  
## $ meta : chr "No" "No" "No" "No" ...  
## $ hema : chr "No" "No" "No" "No" ...  
## $ seps : chr "No" "Yes" "No" "No" ...  
## $ trauma : chr "No" "No" "No" "No" ...  
## $ ortho : chr "No" "No" "No" "No" ...  
## $ adld3p : int 0 NA NA NA NA 0 NA NA NA NA ...  
## $ urin1 : num NA 1437 599 NA 64 ...  
## $ race : chr "white" "white" "white" "white" ...  
## $ income : chr "Under $11k" "Under $11k" "$25-$50k" "$11-$25k" ...  
## $ ptid : int 5 7 9 10 11 12 13 14 16 17 ...  
## $ survtime: int 240 45 317 37 2 201 235 38 134 209 ...

sapply(rhc, function(x) sum(is.na(x)))

## cat1 cat2 ca sadmdte dschdte dthdte lstctdte death   
## 0 4535 0 0 1 2013 0 0   
## cardiohx chfhx dementhx psychhx chrpulhx renalhx liverhx gibledhx   
## 0 0 0 0 0 0 0 0   
## malighx immunhx transhx amihx age sex edu surv2md1   
## 0 0 0 0 0 0 0 0   
## das2d3pc t3d30 dth30 aps1 scoma1 meanbp1 wblc1 hrt1   
## 0 0 0 0 0 80 0 159   
## resp1 temp1 pafi1 alb1 hema1 bili1 crea1 sod1   
## 136 0 0 0 0 0 0 0   
## pot1 paco21 ph1 swang1 wtkilo1 dnr1 ninsclas resp   
## 0 0 0 0 515 0 0 0   
## card neuro gastr renal meta hema seps trauma   
## 0 0 0 0 0 0 0 0   
## ortho adld3p urin1 race income ptid survtime   
## 0 4296 3028 0 0 0 0

Converting the variables that were characters to become factors. Also turning swang1 (indicator if RHC or noRHC into an integer)

rhc1 = rhc %>%   
 mutate(  
 cat1 = as.factor(cat1),  
 cat2 = as.factor(cat2),  
 ca = as.factor(ca),  
 sex = as.factor(sex),  
 sex = as.integer(sex),  
 swang1 = as.factor(swang1),  
 dnr1 = as.factor(dnr1),  
 resp = as.factor(resp),  
 card = as.factor(card),  
 neuro = as.factor(neuro),  
 gastr = as.factor(gastr),  
 renal = as.factor(renal),  
 meta = as.factor(meta),  
 hema = as.factor(hema),  
 seps = as.factor(seps),  
 trauma = as.factor(trauma),  
 ortho = as.factor(ortho),  
 dnr1 = as.integer(dnr1),  
 resp = as.integer(resp),  
 card = as.integer(card),  
 neuro = as.integer(neuro),  
 gastr = as.integer(gastr),  
 renal = as.integer(renal),  
 meta = as.integer(meta),  
 hema = as.integer(hema),  
 seps = as.integer(seps),  
 trauma = as.integer(trauma),  
 ortho = as.integer(ortho),  
 ninsclas = as.factor(ninsclas),  
 race = as.factor(race),  
 income = as.factor(income),  
 swang1 = as.integer(swang1)  
 )  
  
str(rhc1)

## 'data.frame': 5735 obs. of 63 variables:  
## $ cat1 : Factor w/ 9 levels "ARF","CHF","Cirrhosis",..: 6 9 8 1 9 6 8 1 8 1 ...  
## $ cat2 : Factor w/ 6 levels "Cirrhosis","Colon Cancer",..: NA NA 6 NA NA NA NA 3 NA NA ...  
## $ ca : Factor w/ 3 levels "Metastatic","No",..: 3 2 3 2 2 2 1 2 3 3 ...  
## $ sadmdte : int 11142 11799 12083 11146 12035 12389 12381 11453 12426 11381 ...  
## $ dschdte : int 11151 11844 12143 11183 12037 12396 12423 11487 12437 11400 ...  
## $ dthdte : int NA 11844 NA 11183 12037 NA NA 11491 NA NA ...  
## $ lstctdte: int 11382 11844 12400 11182 12036 12590 12616 11490 12560 11590 ...  
## $ death : int 0 1 0 1 1 0 0 1 0 0 ...  
## $ cardiohx: int 0 1 0 0 0 0 0 0 0 0 ...  
## $ chfhx : int 0 1 0 0 0 1 0 0 0 0 ...  
## $ dementhx: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ psychhx : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ chrpulhx: int 1 0 0 0 0 1 0 0 0 0 ...  
## $ renalhx : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ liverhx : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ gibledhx: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ malighx : int 1 0 1 0 0 0 1 0 0 1 ...  
## $ immunhx : int 0 1 1 1 0 0 0 0 0 0 ...  
## $ transhx : int 0 1 0 0 0 0 0 1 0 0 ...  
## $ amihx : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ age : num 70.3 78.2 46.1 75.3 67.9 ...  
## $ sex : int 2 1 1 1 2 1 2 2 1 1 ...  
## $ edu : num 12 12 14.07 9 9.95 ...  
## $ surv2md1: num 0.641 0.755 0.317 0.441 0.437 ...  
## $ das2d3pc: num 23.5 14.8 18.1 22.9 21.1 ...  
## $ t3d30 : int 30 30 30 30 2 30 30 30 30 30 ...  
## $ dth30 : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ aps1 : int 46 50 82 48 72 38 29 25 47 48 ...  
## $ scoma1 : int 0 0 0 0 41 0 26 100 0 0 ...  
## $ meanbp1 : num 41 63 57 55 65 115 67 128 53 73 ...  
## $ wblc1 : num 22.1 28.9 0.05 23.3 29.7 ...  
## $ hrt1 : int 124 137 130 58 125 134 135 102 118 141 ...  
## $ resp1 : num 10 38 40 26 27 36 10 34 30 40 ...  
## $ temp1 : num 38.7 38.9 36.4 35.8 34.8 ...  
## $ pafi1 : num 68 218 276 157 478 ...  
## $ alb1 : num 3.5 2.6 3.5 3.5 3.5 ...  
## $ hema1 : num 58 32.5 21.1 26.3 24 ...  
## $ bili1 : num 1.01 0.7 1.01 0.4 1.01 ...  
## $ crea1 : num 1.2 0.6 2.6 1.7 3.6 ...  
## $ sod1 : int 145 137 146 117 126 138 136 136 136 146 ...  
## $ pot1 : num 4 3.3 2.9 5.8 5.8 ...  
## $ paco21 : num 40 34 16 30 17 68 45 26 40 30 ...  
## $ ph1 : num 7.36 7.33 7.36 7.46 7.23 ...  
## $ swang1 : int 1 2 2 1 2 1 1 1 1 2 ...  
## $ wtkilo1 : num 64.7 45.7 NA 54.6 78.4 ...  
## $ dnr1 : int 1 1 1 1 2 1 1 1 1 1 ...  
## $ ninsclas: Factor w/ 6 levels "Medicaid","Medicare",..: 2 6 5 6 2 2 5 5 5 1 ...  
## $ resp : int 2 1 1 2 1 2 1 2 1 1 ...  
## $ card : int 2 1 2 1 2 1 1 1 1 1 ...  
## $ neuro : int 1 1 1 1 1 1 1 2 1 2 ...  
## $ gastr : int 1 1 1 1 1 1 1 1 1 2 ...  
## $ renal : int 1 1 1 1 1 1 2 1 1 1 ...  
## $ meta : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ hema : int 1 1 1 1 1 1 1 1 2 1 ...  
## $ seps : int 1 2 1 1 1 1 1 2 1 1 ...  
## $ trauma : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ ortho : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ adld3p : int 0 NA NA NA NA 0 NA NA NA NA ...  
## $ urin1 : num NA 1437 599 NA 64 ...  
## $ race : Factor w/ 3 levels "black","other",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ income : Factor w/ 4 levels "$11-$25k","$25-$50k",..: 4 4 2 1 4 4 2 2 4 4 ...  
## $ ptid : int 5 7 9 10 11 12 13 14 16 17 ...  
## $ survtime: int 240 45 317 37 2 201 235 38 134 209 ...

init = mice(rhc1, maxit=0)   
meth = init$method  
predM = init$predictorMatrix

set.seed(103)  
imputed = mice(rhc1, predictorMatrix = predM, m=5)

##   
## iter imp variable  
## 1 1 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 1 2 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 1 3 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 1 4 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 1 5 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 2 1 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 2 2 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 2 3 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 2 4 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 2 5 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 3 1 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 3 2 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 3 3 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 3 4 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 3 5 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 4 1 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 4 2 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 4 3 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 4 4 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 4 5 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 5 1 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 5 2 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 5 3 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 5 4 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1  
## 5 5 cat2 dschdte dthdte meanbp1 hrt1 resp1 wtkilo1 adld3p urin1

## Warning: Number of logged events: 220

fit1 = glm(swang1 ~ (cat1 + ca + sadmdte + dschdte + dthdte + lstctdte + death + cardiohx + chfhx + dementhx + psychhx + chrpulhx + renalhx + liverhx + gibledhx + malighx + immunhx + transhx + amihx + age + sex + edu + surv2md1 + das2d3pc + t3d30 + dth30 + aps1 + scoma1 + wblc1 + temp1 + pafi1 + alb1 + hema1 + bili1 + crea1 + sod1 + pot1 + paco21 + ph1 + dnr1 + ninsclas + resp + card + neuro + gastr + renal + meta + hema + seps + trauma + ortho + adld3p + race + income + ptid + survtime), data = rhc1)

summary(fit1) # display results

##   
## Call:  
## glm(formula = swang1 ~ (cat1 + ca + sadmdte + dschdte + dthdte +   
## lstctdte + death + cardiohx + chfhx + dementhx + psychhx +   
## chrpulhx + renalhx + liverhx + gibledhx + malighx + immunhx +   
## transhx + amihx + age + sex + edu + surv2md1 + das2d3pc +   
## t3d30 + dth30 + aps1 + scoma1 + wblc1 + temp1 + pafi1 + alb1 +   
## hema1 + bili1 + crea1 + sod1 + pot1 + paco21 + ph1 + dnr1 +   
## ninsclas + resp + card + neuro + gastr + renal + meta + hema +   
## seps + trauma + ortho + adld3p + race + income + ptid + survtime),   
## data = rhc1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.81091 -0.26808 -0.09545 0.29223 1.09566   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.667e+00 1.891e+00 1.940 0.05286 .   
## cat1CHF 1.887e-01 6.509e-02 2.898 0.00388 \*\*   
## cat1Cirrhosis -2.526e-01 1.307e-01 -1.932 0.05377 .   
## cat1Colon Cancer 9.593e-02 3.144e-01 0.305 0.76034   
## cat1Coma 5.753e-01 4.393e-01 1.310 0.19077   
## cat1COPD 9.359e-02 7.149e-02 1.309 0.19092   
## cat1Lung Cancer 9.680e-02 1.308e-01 0.740 0.45948   
## cat1MOSF w/Malignancy 3.856e-02 7.308e-02 0.528 0.59795   
## cat1MOSF w/Sepsis 2.829e-01 6.088e-02 4.646 4.10e-06 \*\*\*  
## caNo 3.592e-01 1.881e-01 1.909 0.05665 .   
## caYes 6.390e-02 6.716e-02 0.952 0.34170   
## sadmdte -1.484e-03 1.176e-03 -1.262 0.20753   
## dschdte 1.497e-03 1.161e-03 1.289 0.19793   
## dthdte -2.845e-05 5.924e-05 -0.480 0.63128   
## lstctdte -1.642e-05 1.238e-04 -0.133 0.89448   
## death NA NA NA NA   
## cardiohx 3.003e-03 4.107e-02 0.073 0.94172   
## chfhx 6.872e-02 4.613e-02 1.490 0.13680   
## dementhx -4.668e-02 8.483e-02 -0.550 0.58236   
## psychhx -1.245e-02 6.551e-02 -0.190 0.84929   
## chrpulhx -1.546e-02 4.432e-02 -0.349 0.72727   
## renalhx -2.828e-02 8.420e-02 -0.336 0.73712   
## liverhx -5.672e-02 1.232e-01 -0.460 0.64546   
## gibledhx -4.807e-03 1.172e-01 -0.041 0.96731   
## malighx 1.427e-01 1.649e-01 0.865 0.38723   
## immunhx 4.216e-02 3.587e-02 1.176 0.24021   
## transhx 2.472e-01 5.959e-02 4.148 3.82e-05 \*\*\*  
## amihx 5.470e-02 7.469e-02 0.732 0.46419   
## age -3.048e-03 1.619e-03 -1.883 0.06020 .   
## sex 8.183e-02 3.447e-02 2.374 0.01790 \*   
## edu 2.225e-03 5.793e-03 0.384 0.70108   
## surv2md1 -6.462e-01 2.045e-01 -3.160 0.00165 \*\*   
## das2d3pc -8.644e-03 2.900e-03 -2.981 0.00298 \*\*   
## t3d30 -7.748e-03 4.374e-03 -1.771 0.07697 .   
## dth30 -1.162e-01 7.522e-02 -1.545 0.12289   
## aps1 -1.798e-03 1.369e-03 -1.313 0.18950   
## scoma1 -9.461e-04 1.283e-03 -0.737 0.46122   
## wblc1 1.210e-03 1.153e-03 1.050 0.29427   
## temp1 4.337e-03 1.112e-02 0.390 0.69658   
## pafi1 -1.406e-04 1.669e-04 -0.843 0.39981   
## alb1 -2.410e-02 3.114e-02 -0.774 0.43941   
## hema1 -3.515e-04 2.428e-03 -0.145 0.88495   
## bili1 -6.568e-03 6.082e-03 -1.080 0.28061   
## crea1 -2.039e-03 9.942e-03 -0.205 0.83758   
## sod1 -2.323e-03 2.497e-03 -0.930 0.35258   
## pot1 -3.148e-02 1.740e-02 -1.808 0.07100 .   
## paco21 -3.918e-03 1.961e-03 -1.998 0.04612 \*   
## ph1 -1.126e-01 2.217e-01 -0.508 0.61164   
## dnr1 -7.197e-02 5.522e-02 -1.303 0.19292   
## ninsclasMedicare 5.603e-02 6.454e-02 0.868 0.38559   
## ninsclasMedicare & Medicaid 8.364e-02 8.595e-02 0.973 0.33083   
## ninsclasNo insurance 1.391e-01 9.291e-02 1.497 0.13487   
## ninsclasPrivate 1.419e-01 6.276e-02 2.261 0.02410 \*   
## ninsclasPrivate & Medicare 5.434e-02 6.905e-02 0.787 0.43156   
## resp -3.283e-02 4.377e-02 -0.750 0.45356   
## card 1.179e-01 4.350e-02 2.711 0.00689 \*\*   
## neuro -9.634e-02 8.713e-02 -1.106 0.26931   
## gastr 8.221e-02 6.594e-02 1.247 0.21296   
## renal 3.536e-03 8.145e-02 0.043 0.96539   
## meta -3.467e-02 7.586e-02 -0.457 0.64786   
## hema -9.495e-02 6.821e-02 -1.392 0.16436   
## seps 4.997e-02 5.533e-02 0.903 0.36679   
## trauma -1.177e-01 3.024e-01 -0.389 0.69723   
## ortho NA NA NA NA   
## adld3p -2.014e-02 9.544e-03 -2.110 0.03523 \*   
## raceother 4.298e-02 7.365e-02 0.584 0.55971   
## racewhite -3.432e-02 4.727e-02 -0.726 0.46804   
## income$25-$50k 6.378e-03 5.311e-02 0.120 0.90444   
## income> $50k 3.242e-02 6.937e-02 0.467 0.64038   
## incomeUnder $11k -3.590e-03 4.233e-02 -0.085 0.93244   
## ptid 1.648e-06 5.225e-06 0.315 0.75260   
## survtime NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.1587574)  
##   
## Null deviance: 140.69 on 705 degrees of freedom  
## Residual deviance: 101.13 on 637 degrees of freedom  
## (5029 observations deleted due to missingness)  
## AIC: 771.63  
##   
## Number of Fisher Scoring iterations: 2

confint(fit1) # 95% CI for the coefficients

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -3.827972e-02 7.372938e+00  
## cat1CHF 6.108355e-02 3.162423e-01  
## cat1Cirrhosis -5.087955e-01 3.621660e-03  
## cat1Colon Cancer -5.202095e-01 7.120778e-01  
## cat1Coma -2.856421e-01 1.436277e+00  
## cat1COPD -4.651595e-02 2.337017e-01  
## cat1Lung Cancer -1.595304e-01 3.531290e-01  
## cat1MOSF w/Malignancy -1.046695e-01 1.817799e-01  
## cat1MOSF w/Sepsis 1.635519e-01 4.021921e-01  
## caNo -9.498020e-03 7.278524e-01  
## caYes -6.772608e-02 1.955342e-01  
## sadmdte -3.788630e-03 8.211853e-04  
## dschdte -7.793748e-04 3.772750e-03  
## dthdte -1.445624e-04 8.766951e-05  
## lstctdte -2.589798e-04 2.261375e-04  
## death NA NA  
## cardiohx -7.748631e-02 8.349304e-02  
## chfhx -2.169291e-02 1.591295e-01  
## dementhx -2.129490e-01 1.195941e-01  
## psychhx -1.408465e-01 1.159407e-01  
## chrpulhx -1.023327e-01 7.140390e-02  
## renalhx -1.933078e-01 1.367554e-01  
## liverhx -2.982311e-01 1.847928e-01  
## gibledhx -2.345883e-01 2.249739e-01  
## malighx -1.804968e-01 4.658158e-01  
## immunhx -2.813231e-02 1.124564e-01  
## transhx 1.303574e-01 3.639458e-01  
## amihx -9.168234e-02 2.010844e-01  
## age -6.221740e-03 1.251586e-04  
## sex 1.426389e-02 1.493864e-01  
## edu -9.129296e-03 1.357880e-02  
## surv2md1 -1.046970e+00 -2.453525e-01  
## das2d3pc -1.432691e-02 -2.960435e-03  
## t3d30 -1.632102e-02 8.248194e-04  
## dth30 -2.636104e-01 3.122702e-02  
## aps1 -4.481638e-03 8.851142e-04  
## scoma1 -3.461138e-03 1.568939e-03  
## wblc1 -1.049317e-03 3.469261e-03  
## temp1 -1.745458e-02 2.612957e-02  
## pafi1 -4.676680e-04 1.864715e-04  
## alb1 -8.513753e-02 3.694635e-02  
## hema1 -5.110390e-03 4.407433e-03  
## bili1 -1.848917e-02 5.353001e-03  
## crea1 -2.152589e-02 1.744797e-02  
## sod1 -7.216999e-03 2.571169e-03  
## pot1 -6.558642e-02 2.636369e-03  
## paco21 -7.760673e-03 -7.502509e-05  
## ph1 -5.471219e-01 3.218916e-01  
## dnr1 -1.801944e-01 3.625539e-02  
## ninsclasMedicare -7.045580e-02 1.825214e-01  
## ninsclasMedicare & Medicaid -8.481379e-02 2.521023e-01  
## ninsclasNo insurance -4.300679e-02 3.211766e-01  
## ninsclasPrivate 1.889348e-02 2.649228e-01  
## ninsclasPrivate & Medicare -8.098928e-02 1.896752e-01  
## resp -1.186241e-01 5.296746e-02  
## card 3.266686e-02 2.031701e-01  
## neuro -2.671194e-01 7.444258e-02  
## gastr -4.703183e-02 2.114507e-01  
## renal -1.561042e-01 1.631755e-01  
## meta -1.833586e-01 1.140260e-01  
## hema -2.286364e-01 3.872729e-02  
## seps -5.847206e-02 1.584105e-01  
## trauma -7.104003e-01 4.749896e-01  
## ortho NA NA  
## adld3p -3.884434e-02 -1.433450e-03  
## raceother -1.013649e-01 1.873200e-01  
## racewhite -1.269700e-01 5.832441e-02  
## income$25-$50k -9.771348e-02 1.104700e-01  
## income> $50k -1.035457e-01 1.683948e-01  
## incomeUnder $11k -8.656432e-02 7.938359e-02  
## ptid -8.593624e-06 1.188916e-05  
## survtime NA NA

exp(coef(fit1)) # exponentiated coefficients

## (Intercept) cat1CHF   
## 39.1472007 1.2076338   
## cat1Cirrhosis cat1Colon Cancer   
## 0.7767887 1.1006865   
## cat1Coma cat1COPD   
## 1.7776947 1.0981126   
## cat1Lung Cancer cat1MOSF w/Malignancy   
## 1.1016392 1.0393081   
## cat1MOSF w/Sepsis caNo   
## 1.3269353 1.4321505   
## caYes sadmdte   
## 1.0659901 0.9985174   
## dschdte dthdte   
## 1.0014978 0.9999716   
## lstctdte death   
## 0.9999836 NA   
## cardiohx chfhx   
## 1.0030079 1.0711344   
## dementhx psychhx   
## 0.9543952 0.9876243   
## chrpulhx renalhx   
## 0.9846545 0.9721198   
## liverhx gibledhx   
## 0.9448594 0.9952043   
## malighx immunhx   
## 1.1533371 1.0430635   
## transhx amihx   
## 1.2803732 1.0562248   
## age sex   
## 0.9969564 1.0852660   
## edu surv2md1   
## 1.0022272 0.5240537   
## das2d3pc t3d30   
## 0.9913936 0.9922818   
## dth30 aps1   
## 0.8903045 0.9982034   
## scoma1 wblc1   
## 0.9990543 1.0012107   
## temp1 pafi1   
## 1.0043469 0.9998594   
## alb1 hema1   
## 0.9761924 0.9996486   
## bili1 crea1   
## 0.9934534 0.9979631   
## sod1 pot1   
## 0.9976798 0.9690152   
## paco21 ph1   
## 0.9960898 0.8934945   
## dnr1 ninsclasMedicare   
## 0.9305593 1.0576324   
## ninsclasMedicare & Medicaid ninsclasNo insurance   
## 1.0872421 1.1492217   
## ninsclasPrivate ninsclasPrivate & Medicare   
## 1.1524708 1.0558467   
## resp card   
## 0.9677047 1.1251524   
## neuro gastr   
## 0.9081566 1.0856831   
## renal meta   
## 1.0035419 0.9659277   
## hema seps   
## 0.9094143 1.0512387   
## trauma ortho   
## 0.8889579 NA   
## adld3p raceother   
## 0.9800625 1.0439145   
## racewhite income$25-$50k   
## 0.9662596 1.0063986   
## income> $50k incomeUnder $11k   
## 1.0329560 0.9964161   
## ptid survtime   
## 1.0000016 NA

exp(confint(fit1)) # 95% CI for exponentiated coefficients

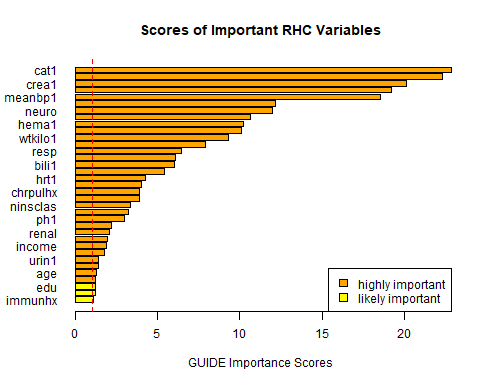
## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.9624437 1592.3044049  
## cat1CHF 1.0629877 1.3719627  
## cat1Cirrhosis 0.6012193 1.0036282  
## cat1Colon Cancer 0.5943960 2.0382218  
## cat1Coma 0.7515315 4.2050111  
## cat1COPD 0.9545493 1.2632676  
## cat1Lung Cancer 0.8525441 1.4235147  
## cat1MOSF w/Malignancy 0.9006221 1.1993502  
## cat1MOSF w/Sepsis 1.1776865 1.4950985  
## caNo 0.9905469 2.0706290  
## caYes 0.9345164 1.2159604  
## sadmdte 0.9962185 1.0008215  
## dschdte 0.9992209 1.0037799  
## dthdte 0.9998554 1.0000877  
## lstctdte 0.9997411 1.0002262  
## death NA NA  
## cardiohx 0.9254397 1.0870777  
## chfhx 0.9785407 1.1724898  
## dementhx 0.8081974 1.1270393  
## psychhx 0.8686226 1.1229293  
## chrpulhx 0.9027291 1.0740149  
## renalhx 0.8242282 1.1465477  
## liverhx 0.7421298 1.2029691  
## gibledhx 0.7908964 1.2522901  
## malighx 0.8348554 1.5933136  
## immunhx 0.9722597 1.1190235  
## transhx 1.1392354 1.4389962  
## amihx 0.9123949 1.2227280  
## age 0.9937976 1.0001252  
## sex 1.0143661 1.1611216  
## edu 0.9909122 1.0136714  
## surv2md1 0.3509998 0.7824287  
## das2d3pc 0.9857752 0.9970439  
## t3d30 0.9838115 1.0008252  
## dth30 0.7682728 1.0317197  
## aps1 0.9955284 1.0008855  
## scoma1 0.9965448 1.0015702  
## wblc1 0.9989512 1.0034753  
## temp1 0.9826969 1.0264739  
## pafi1 0.9995324 1.0001865  
## alb1 0.9183860 1.0376374  
## hema1 0.9949026 1.0044172  
## bili1 0.9816807 1.0053674  
## crea1 0.9787041 1.0176011  
## sod1 0.9928090 1.0025745  
## pot1 0.9365181 1.0026398  
## paco21 0.9922694 0.9999250  
## ph1 0.5786127 1.3797352  
## dnr1 0.8351078 1.0369206  
## ninsclasMedicare 0.9319689 1.2002398  
## ninsclasMedicare & Medicaid 0.9186833 1.2867277  
## ninsclasNo insurance 0.9579049 1.3787491  
## ninsclasPrivate 1.0190731 1.3033303  
## ninsclasPrivate & Medicare 0.9222036 1.2088569  
## resp 0.8881416 1.0543953  
## card 1.0332063 1.2252809  
## neuro 0.7655816 1.0772835  
## gastr 0.9540570 1.2354690  
## renal 0.8554701 1.1772433  
## meta 0.8324696 1.1207812  
## hema 0.7956178 1.0394870  
## seps 0.9432046 1.1716471  
## trauma 0.4914474 1.6079975  
## ortho NA NA  
## adld3p 0.9619004 0.9985676  
## raceother 0.9036033 1.2060132  
## racewhite 0.8807601 1.0600588  
## income$25-$50k 0.9069087 1.1168028  
## income> $50k 0.9016348 1.1834038  
## incomeUnder $11k 0.9170766 1.0826195  
## ptid 0.9999914 1.0000119  
## survtime NA NA

pred = predict(fit1, type="response") # predicted values  
res = residuals(fit1, type="deviance") # predicted residuals

# Part 2 -- Plotting Important Scores

par(las = 1, mar = c(5,5,4,2), cex = .75)  
  
leg.col= c("orange","yellow")  
leg.txt = c("highly important","likely important")  
  
x = read.table("imp.scr",header=TRUE)  
score = x$Score  
vars = x$Variable  
type = x$Type  
  
barcol = rep("orange",length(vars))  
barcol[type == "L"] = "yellow"  
barcol[type == "U"] = "cyan"  
n = sum(x$Type != "U")  
  
barplot(rev(score[1:n]), names.arg = rev(vars[1:n]), col = rev(barcol[1:n]), horiz = T, xlab = "GUIDE Importance Scores", main = "Scores of Important RHC Variables")  
  
abline(v = 1, col = "red", lty = 2)  
legend("bottomright", legend = leg.txt, fill = leg.col)



# Part 3

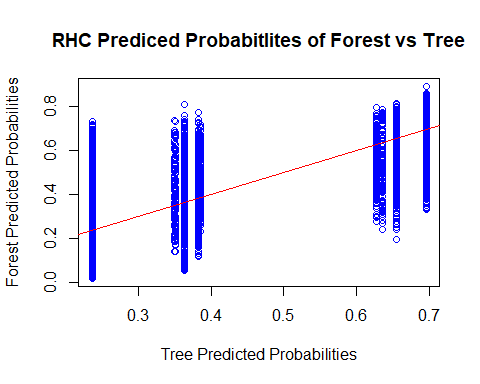
The logistic model isn’t as good as the GUIDE model. It does not determine that as many variables are statistically significantly, and highly important as the GUIDE model does. Furthermore, the logistic model looks at the individual types within each variable instead of the overall significance of the variable.

# Part 4

The GUIDE method is much more accurate because it did not need to convert the types of the variables, but was able to process them as is. So for variables with several factors, guide was able to intelligently interpret this whereas R ranked them based off whatever came first.

R code for Plotting Predicted Probabilities

tree = read.table("classpred.txt",header=TRUE)  
forest = read.table("forestpred.txt",header=TRUE)  
tree.p = tree[,6]  
forest.p = forest[,3]  
plot(forest.p ~ tree.p, xlab="Tree Predicted Probabilities",  
ylab="Forest Predicted Probabilities",col="blue", main = "RHC Prediced Probabitlites of Forest vs Tree")  
abline(c(0,1),col="red")



# Part 5

Attached to my canvas submission is my input and output files for the GUIDE forest (forestin.txt and forestout.txt) and important scores (impin.txt and impout.txt).

