Outcome Calculator After Head Traumas

Anthony Morada

9/18/2020

Table of Contents

# Setting up and Cleaning Data

knitr::opts\_chunk$set(  
 echo = TRUE,  
 message = FALSE,  
 warning = FALSE  
)  
#Data is set up on two sheets. Therefore, we will load both and merge the data  
library(readxl)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

library(rcompanion)  
library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library(InformationValue)

##   
## Attaching package: 'InformationValue'

## The following objects are masked from 'package:caret':  
##   
## confusionMatrix, precision, sensitivity, specificity

sheet1 <- read\_excel("~/THE GUTHRIE CLINIC/Foundation Research Projects - GCS Trauma Calculator/Master Doc.xlsx",  
 sheet = "Sheet1",  
 col\_types = c("text", "numeric", "text",   
 "text", "text", "text", "numeric",   
 "text", "text", "text", "text", "text",   
 "text", "text", "text", "numeric",   
 "numeric"))  
sheet2 <- read\_excel("~/THE GUTHRIE CLINIC/Foundation Research Projects - GCS Trauma Calculator/Master Doc.xlsx",  
 sheet = "Sheet2",  
 col\_types = c("text", "numeric", "text",   
 "text", "text", "text", "numeric",   
 "text", "text", "text", "text", "text",   
 "text", "text", "text", "numeric",   
 "numeric"))  
  
colnames(sheet2) <- colnames(sheet1)  
df <- rbind(sheet1, sheet2)  
  
rm(sheet1, sheet2)

df$`PEG Tube` <- !is.na(df$`PEG Tube`)  
df$Tracheostomy <- !is.na(df$Tracheostomy)  
df$Alcohol <- as.factor(df$Alcohol)  
df$Stroke <- grepl("Stroke / CVA", df$Stroke)  
df$`Severe Sepsis` <- grepl("Severe sepsis", df$`Severe Sepsis`)  
df$`Pulmonary Embolism` <- grepl("Pulmonary embolism", df$`Pulmonary Embolism`)  
df$Pneumonia <-gsub("0",NA,df$Pneumonia)  
df$Pneumonia<- grepl("Pneumonia", df$Pneumonia, ignore.case = TRUE)  
df$`Myocardial infarction` <- grepl("Myocardial infarction", df$`Myocardial infarction`)  
df$DVT <- grepl("DVT", df$DVT, ignore.case = TRUE)  
df$`Cardiac arrest` <- grepl("Cardiac arrest", df$`Cardiac arrest`)  
df$Gender <- as.factor(df$Gender)  
df <- select(df, -inc\_key)  
  
#Outcomes  
df$Expired <- grepl("expired", df$Discharge, ignore.case = TRUE)  
df$Home <- grepl("home", df$Discharge, ignore.case = TRUE)  
df$Other <- grepl("psychiatric hospital|short-term general hospital|Intermediate Care Facility|another type of institution|rehabilitation or long term|hospice care|inpatient rehab|Long Term Care Hospital|Skilled Nursing Facility", df$Discharge)

# Expiration Model

#Check Class bias  
table(df$Expired)

##   
## FALSE TRUE   
## 1082459 57424

# Create Training Data  
input\_ones <- df[which(df$Expired == TRUE), ] # all 1's  
input\_zeros <- df[which(df$Expired == FALSE), ] # all 0's  
set.seed(100) # for repeatability of samples  
input\_ones\_training\_rows <- sample(1:nrow(input\_ones), 0.7\*nrow(input\_ones)) # 1's for training  
input\_zeros\_training\_rows <- sample(1:nrow(input\_zeros), 0.7\*nrow(input\_ones)) # 0's for training. Pick as many 0's as 1's  
training\_ones <- input\_ones[input\_ones\_training\_rows, ]   
training\_zeros <- input\_zeros[input\_zeros\_training\_rows, ]  
trainingData <- rbind(training\_ones, training\_zeros) # row bind the 1's and 0's   
  
# Create Test Data  
test\_ones <- input\_ones[-input\_ones\_training\_rows, ]  
test\_zeros <- input\_zeros[-input\_zeros\_training\_rows, ]  
testData <- rbind(test\_ones, test\_zeros) # row bind the 1's and 0's   
#Determining model with step procedure  
### Define full and null models and do step procedure  
  
model.full = glm(Expired ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
 data=trainingData,   
 family=binomial(link="logit"))  
  
step(model.full,  
 scope = list(upper=model.full),  
 direction="both",  
 test="Chisq",  
 data=trainingData)

## Start: AIC=61321.76  
## Expired ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
##   
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 61300 61322   
## - Alcohol 5 61477 61489 177.4 < 2.2e-16 \*\*\*  
## - Gender 2 61613 61631 313.0 < 2.2e-16 \*\*\*  
## - Age 1 68714 68734 7414.5 < 2.2e-16 \*\*\*  
## - ISSAIS 1 69543 69563 8243.6 < 2.2e-16 \*\*\*  
## - GCSTOT 1 79373 79393 18072.9 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Call: glm(formula = Expired ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
## family = binomial(link = "logit"), data = trainingData)  
##   
## Coefficients:  
## (Intercept)   
## -1.221497   
## Age   
## 0.044955   
## GenderMale   
## 0.393186   
## GenderNot Known/Not Recorded BIU 2   
## 0.527451   
## ISSAIS   
## 0.096683   
## AlcoholNo (not tested)   
## 0.013967   
## AlcoholNot Applicable BIU 1   
## -0.035164   
## AlcoholNot Known/Not Recorded BIU 2   
## -0.007105   
## AlcoholYes (confirmed by test [beyond legal limit])   
## -0.431021   
## AlcoholYes (confirmed by test [trace levels])   
## 0.066349   
## GCSTOT   
## -0.299598   
##   
## Degrees of Freedom: 80391 Total (i.e. Null); 80381 Residual  
## Null Deviance: 111400   
## Residual Deviance: 61300 AIC: 61320

#Final model  
model.final = glm(Expired ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,  
 data=trainingData,  
 family = binomial(link="logit"),  
 na.action(na.omit)  
 )  
  
summary(model.final)

##   
## Call:  
## glm(formula = Expired ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
## family = binomial(link = "logit"), data = trainingData, weights = na.action(na.omit))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -4.3647 -0.6030 -0.0508 0.4795 3.0099   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) -1.2214973 0.0551715  
## Age 0.0449550 0.0005803  
## GenderMale 0.3931863 0.0223097  
## GenderNot Known/Not Recorded BIU 2 0.5274508 0.6094410  
## ISSAIS 0.0966827 0.0012136  
## AlcoholNo (not tested) 0.0139673 0.0239892  
## AlcoholNot Applicable BIU 1 -0.0351640 0.0605434  
## AlcoholNot Known/Not Recorded BIU 2 -0.0071048 0.0573911  
## AlcoholYes (confirmed by test [beyond legal limit]) -0.4310212 0.0355099  
## AlcoholYes (confirmed by test [trace levels]) 0.0663486 0.0535447  
## GCSTOT -0.2995983 0.0026502  
## z value Pr(>|z|)   
## (Intercept) -22.140 <2e-16 \*\*\*  
## Age 77.465 <2e-16 \*\*\*  
## GenderMale 17.624 <2e-16 \*\*\*  
## GenderNot Known/Not Recorded BIU 2 0.865 0.387   
## ISSAIS 79.664 <2e-16 \*\*\*  
## AlcoholNo (not tested) 0.582 0.560   
## AlcoholNot Applicable BIU 1 -0.581 0.561   
## AlcoholNot Known/Not Recorded BIU 2 -0.124 0.901   
## AlcoholYes (confirmed by test [beyond legal limit]) -12.138 <2e-16 \*\*\*  
## AlcoholYes (confirmed by test [trace levels]) 1.239 0.215   
## GCSTOT -113.047 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 111447 on 80391 degrees of freedom  
## Residual deviance: 61300 on 80381 degrees of freedom  
## AIC: 61322  
##   
## Number of Fisher Scoring iterations: 5

print(exp(cbind(OR = coef(model.final), confint(model.final))))

## OR 2.5 %  
## (Intercept) 0.2947885 0.2645629  
## Age 1.0459808 1.0447953  
## GenderMale 1.4816944 1.4183305  
## GenderNot Known/Not Recorded BIU 2 1.6946068 0.5146241  
## ISSAIS 1.1015108 1.0989017  
## AlcoholNo (not tested) 1.0140653 0.9675063  
## AlcoholNot Applicable BIU 1 0.9654470 0.8571061  
## AlcoholNot Known/Not Recorded BIU 2 0.9929204 0.8870270  
## AlcoholYes (confirmed by test [beyond legal limit]) 0.6498452 0.6061199  
## AlcoholYes (confirmed by test [trace levels]) 1.0685992 0.9620845  
## GCSTOT 0.7411158 0.7372617  
## 97.5 %  
## (Intercept) 0.3284436  
## Age 1.0471749  
## GenderMale 1.5479602  
## GenderNot Known/Not Recorded BIU 2 5.5590696  
## ISSAIS 1.1041425  
## AlcoholNo (not tested) 1.0629078  
## AlcoholNot Applicable BIU 1 1.0867137  
## AlcoholNot Known/Not Recorded BIU 2 1.1108329  
## AlcoholYes (confirmed by test [beyond legal limit]) 0.6966516  
## AlcoholYes (confirmed by test [trace levels]) 1.1867909  
## GCSTOT 0.7449616

#Analysis of variance for individual terms  
Anova(model.final, type="II", test="Wald")

## Analysis of Deviance Table (Type II tests)  
##   
## Response: Expired  
## Df Chisq Pr(>Chisq)   
## Age 1 6000.82 < 2.2e-16 \*\*\*  
## Gender 2 310.81 < 2.2e-16 \*\*\*  
## ISSAIS 1 6346.35 < 2.2e-16 \*\*\*  
## Alcohol 5 175.08 < 2.2e-16 \*\*\*  
## GCSTOT 1 12779.72 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Pseudo-R-squared  
nagelkerke(model.final)

## $Models  
##   
## Model: "glm, Expired ~ Age + Gender + ISSAIS + Alcohol + GCSTOT, binomial(link = \"logit\"), trainingData, na.action(na.omit)"  
## Null: "glm, Expired ~ 1, binomial(link = \"logit\"), trainingData, na.action(na.omit)"   
##   
## $Pseudo.R.squared.for.model.vs.null  
## Pseudo.R.squared  
## McFadden 0.449965  
## Cox and Snell (ML) 0.464087  
## Nagelkerke (Cragg and Uhler) 0.618783  
##   
## $Likelihood.ratio.test  
## Df.diff LogLik.diff Chisq p.value  
## -10 -25074 50147 0  
##   
## $Number.of.observations  
##   
## Model: 80392  
## Null: 80392  
##   
## $Messages  
## [1] "Note: For models fit with REML, these statistics are based on refitting with ML"  
##   
## $Warnings  
## [1] "None"

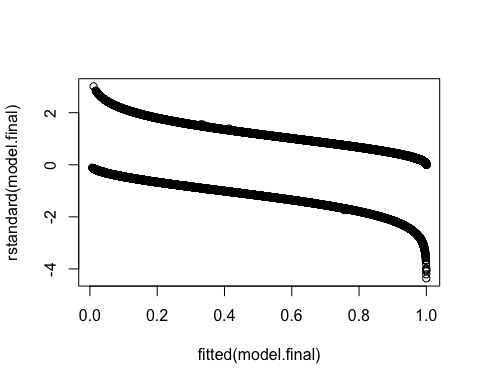
# Overall p-value for model  
### Define null models and compare to final model  
model.null = glm(Expired ~ 1,  
 data=trainingData,  
 family = binomial(link="logit")  
 )  
  
anova(model.final,  
 model.null,  
 test="Chisq")

## Analysis of Deviance Table  
##   
## Model 1: Expired ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
## Model 2: Expired ~ 1  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 80381 61300   
## 2 80391 111447 -10 -50147 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

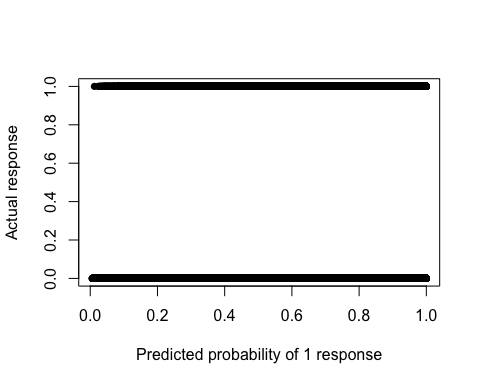
lrtest(model.final)

## Likelihood ratio test  
##   
## Model 1: Expired ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
## Model 2: Expired ~ 1  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 -30650   
## 2 1 -55723 -10 50147 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Plot of standardized residuals  
plot(fitted(model.final),  
 rstandard(model.final))



#prediction  
predicted <- predict(model.final, testData, type = "response")  
preds <- predict(model.final, newdata = testData, type = "response", se.fit = TRUE)  
critval <- 1.96 ## approx 95% CI  
upr <- preds$fit + (critval \* preds$se.fit)  
lwr <- preds$fit - (critval \* preds$se.fit)  
fit <- preds$fit  
  
  
#Simple plot of predicted values  
testData$predy = predict(model.final, testData, type = "response")  
### Plot  
plot(testData$Expired ~ predicted,  
 pch = 16,  
 xlab="Predicted probability of 1 response",  
 ylab="Actual response")



#validate - confusion matrix  
confmatrix <- table(Actual\_Value=testData$Expired, Predicted\_Value=predicted > 0.5)  
confmatrix

## Predicted\_Value  
## Actual\_Value FALSE TRUE  
## FALSE 915090 127173  
## TRUE 3555 13673

#sensitivity  
print("SENSITIVITY")

## [1] "SENSITIVITY"

sensitivity(testData$Expired, predicted)

## [1] 0.7936499

#specificity  
print("SPECIFICITY")

## [1] "SPECIFICITY"

specificity(testData$Expired, predicted)

## [1] 0.8779838

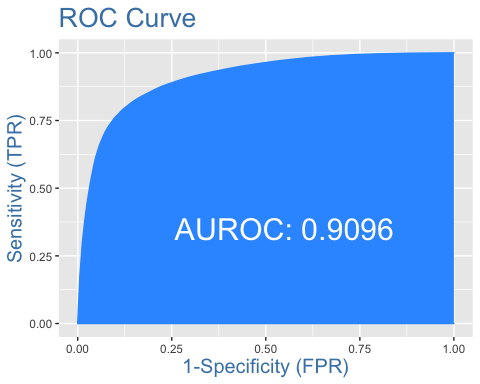
#accuracy  
print("ACCURACY")

## [1] "ACCURACY"

(confmatrix[[1,1]] + confmatrix[[2,2]])/sum(confmatrix)

## [1] 0.8766124

plotROC(testData$Expired, predicted)



#Check for overdispersion  
#One guideline is that if the ratio of the residual deviance to the residual degrees of freedom exceeds 1.5, then the model is overdispersed. Overdispersion indicates that the model doesn’t fit the data well: the explanatory variables may not well describe the dependent variable or the model may not be specified correctly for these data.  
summary(model.final)$deviance / summary(model.final)$df.residual

## [1] 0.7626151

saveRDS(model.final, "expired\_model.rds")

# Disposition to Home Model

#Check Class bias  
table(df$Home)

##   
## FALSE TRUE   
## 461693 678190

# Create Training Data  
input\_ones <- df[which(df$Home == TRUE), ] # all 1's  
input\_zeros <- df[which(df$Home == FALSE), ] # all 0's  
set.seed(100) # for repeatability of samples  
# Because there were more Trues than False, use 0s to set sample  
input\_ones\_training\_rows <- sample(1:nrow(input\_ones), 0.7\*nrow(input\_zeros)) # 1's for training  
input\_zeros\_training\_rows <- sample(1:nrow(input\_zeros), 0.7\*nrow(input\_zeros)) # 0's for training. Pick as many 0's as 1's  
training\_ones <- input\_ones[input\_ones\_training\_rows, ]   
training\_zeros <- input\_zeros[input\_zeros\_training\_rows, ]  
trainingData <- rbind(training\_ones, training\_zeros) # row bind the 1's and 0's   
  
# Create Test Data  
test\_ones <- input\_ones[-input\_ones\_training\_rows, ]  
test\_zeros <- input\_zeros[-input\_zeros\_training\_rows, ]  
testData <- rbind(test\_ones, test\_zeros) # row bind the 1's and 0's   
  
  
#Determining model with step procedure  
### Define full and null models and do step procedure  
model.full = glm(Home ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
 data=trainingData,   
 family=binomial(link="logit"))  
  
step(model.full,  
 scope = list(upper=model.full),  
 direction="both",  
 test="Chisq",  
 data=trainingData)

## Start: AIC=712045.1  
## Home ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
##   
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 712023 712045   
## - Alcohol 5 713057 713069 1034 < 2.2e-16 \*\*\*  
## - Gender 2 715120 715138 3097 < 2.2e-16 \*\*\*  
## - ISSAIS 1 738382 738402 26359 < 2.2e-16 \*\*\*  
## - GCSTOT 1 744013 744033 31990 < 2.2e-16 \*\*\*  
## - Age 1 814443 814463 102420 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Call: glm(formula = Home ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
## family = binomial(link = "logit"), data = trainingData)  
##   
## Coefficients:  
## (Intercept)   
## 0.829704   
## Age   
## -0.046740   
## GenderMale   
## 0.344945   
## GenderNot Known/Not Recorded BIU 2   
## 0.096233   
## ISSAIS   
## -0.063639   
## AlcoholNo (not tested)   
## 0.072823   
## AlcoholNot Applicable BIU 1   
## 0.007997   
## AlcoholNot Known/Not Recorded BIU 2   
## -0.066228   
## AlcoholYes (confirmed by test [beyond legal limit])   
## 0.303447   
## AlcoholYes (confirmed by test [trace levels])   
## 0.075260   
## GCSTOT   
## 0.173256   
##   
## Degrees of Freedom: 646369 Total (i.e. Null); 646359 Residual  
## Null Deviance: 896100   
## Residual Deviance: 712000 AIC: 712000

#Final model  
model.final = glm(Home ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,  
 data=trainingData,  
 family = binomial(link="logit"),  
 na.action(na.omit)  
 )  
  
summary(model.final)

##   
## Call:  
## glm(formula = Home ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
## family = binomial(link = "logit"), data = trainingData, weights = na.action(na.omit))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4979 -0.8964 0.1241 0.9028 3.3762   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 0.8297038 0.0198185  
## Age -0.0467398 0.0001613  
## GenderMale 0.3449449 0.0061932  
## GenderNot Known/Not Recorded BIU 2 0.0962326 0.1604760  
## ISSAIS -0.0636387 0.0004159  
## AlcoholNo (not tested) 0.0728235 0.0069672  
## AlcoholNot Applicable BIU 1 0.0079975 0.0169145  
## AlcoholNot Known/Not Recorded BIU 2 -0.0662276 0.0162514  
## AlcoholYes (confirmed by test [beyond legal limit]) 0.3034473 0.0099539  
## AlcoholYes (confirmed by test [trace levels]) 0.0752596 0.0153370  
## GCSTOT 0.1732565 0.0010424  
## z value Pr(>|z|)   
## (Intercept) 41.865 < 2e-16 \*\*\*  
## Age -289.806 < 2e-16 \*\*\*  
## GenderMale 55.698 < 2e-16 \*\*\*  
## GenderNot Known/Not Recorded BIU 2 0.600 0.549   
## ISSAIS -153.021 < 2e-16 \*\*\*  
## AlcoholNo (not tested) 10.452 < 2e-16 \*\*\*  
## AlcoholNot Applicable BIU 1 0.473 0.636   
## AlcoholNot Known/Not Recorded BIU 2 -4.075 4.60e-05 \*\*\*  
## AlcoholYes (confirmed by test [beyond legal limit]) 30.485 < 2e-16 \*\*\*  
## AlcoholYes (confirmed by test [trace levels]) 4.907 9.25e-07 \*\*\*  
## GCSTOT 166.205 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 896059 on 646369 degrees of freedom  
## Residual deviance: 712023 on 646359 degrees of freedom  
## AIC: 712045  
##   
## Number of Fisher Scoring iterations: 4

print(exp(cbind(OR = coef(model.final), confint(model.final))))

## OR 2.5 %  
## (Intercept) 2.2926395 2.2052698  
## Age 0.9543357 0.9540338  
## GenderMale 1.4119121 1.3948771  
## GenderNot Known/Not Recorded BIU 2 1.1010151 0.8028465  
## ISSAIS 0.9383440 0.9375788  
## AlcoholNo (not tested) 1.0755407 1.0609550  
## AlcoholNot Applicable BIU 1 1.0080295 0.9751478  
## AlcoholNot Known/Not Recorded BIU 2 0.9359179 0.9065719  
## AlcoholYes (confirmed by test [beyond legal limit]) 1.3545202 1.3283592  
## AlcoholYes (confirmed by test [trace levels]) 1.0781640 1.0462514  
## GCSTOT 1.1891711 1.1867466  
## 97.5 %  
## (Intercept) 2.3834275  
## Age 0.9546372  
## GenderMale 1.4291550  
## GenderNot Known/Not Recorded BIU 2 1.5070697  
## ISSAIS 0.9391085  
## AlcoholNo (not tested) 1.0903303  
## AlcoholNot Applicable BIU 1 1.0419966  
## AlcoholNot Known/Not Recorded BIU 2 0.9662047  
## AlcoholYes (confirmed by test [beyond legal limit]) 1.3812152  
## AlcoholYes (confirmed by test [trace levels]) 1.1110828  
## GCSTOT 1.1916061

#Analysis of variance for individual terms  
Anova(model.final, type="II", test="Wald")

## Analysis of Deviance Table (Type II tests)  
##   
## Response: Home  
## Df Chisq Pr(>Chisq)   
## Age 1 83987.4 < 2.2e-16 \*\*\*  
## Gender 2 3102.7 < 2.2e-16 \*\*\*  
## ISSAIS 1 23415.6 < 2.2e-16 \*\*\*  
## Alcohol 5 1023.8 < 2.2e-16 \*\*\*  
## GCSTOT 1 27624.1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Pseudo-R-squared  
nagelkerke(model.final)

## $Models  
##   
## Model: "glm, Home ~ Age + Gender + ISSAIS + Alcohol + GCSTOT, binomial(link = \"logit\"), trainingData, na.action(na.omit)"  
## Null: "glm, Home ~ 1, binomial(link = \"logit\"), trainingData, na.action(na.omit)"   
##   
## $Pseudo.R.squared.for.model.vs.null  
## Pseudo.R.squared  
## McFadden 0.205384  
## Cox and Snell (ML) 0.247777  
## Nagelkerke (Cragg and Uhler) 0.330369  
##   
## $Likelihood.ratio.test  
## Df.diff LogLik.diff Chisq p.value  
## -10 -92018 184040 0  
##   
## $Number.of.observations  
##   
## Model: 646370  
## Null: 646370  
##   
## $Messages  
## [1] "Note: For models fit with REML, these statistics are based on refitting with ML"  
##   
## $Warnings  
## [1] "None"

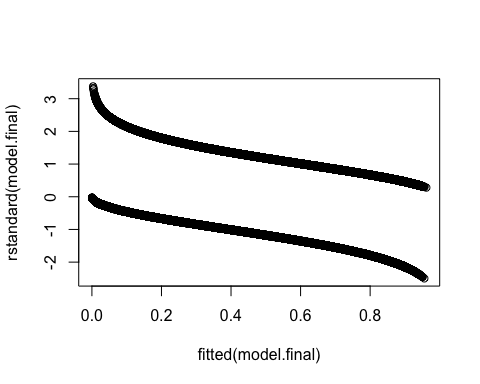
# Overall p-value for model  
### Define null models and compare to final model  
model.null = glm(Home ~ 1,  
 data=trainingData,  
 family = binomial(link="logit")  
 )  
  
anova(model.final,  
 model.null,  
 test="Chisq")

## Analysis of Deviance Table  
##   
## Model 1: Home ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
## Model 2: Home ~ 1  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 646359 712023   
## 2 646369 896059 -10 -184036 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

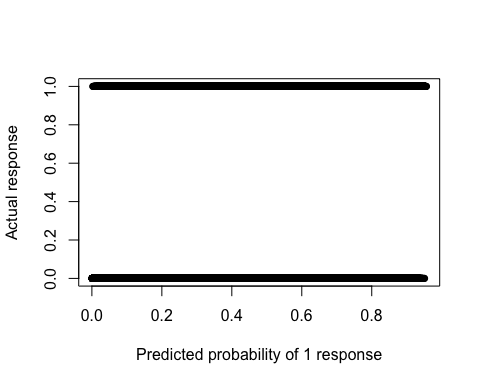
lrtest(model.final)

## Likelihood ratio test  
##   
## Model 1: Home ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
## Model 2: Home ~ 1  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 -356012   
## 2 1 -448030 -10 184036 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Plot of standardized residuals  
plot(fitted(model.final),  
 rstandard(model.final))



#prediction  
predicted <- predict(model.final, testData, type = "response")  
preds <- predict(model.final, newdata = testData, type = "response", se.fit = TRUE)  
critval <- 1.96 ## approx 95% CI  
upr <- preds$fit + (critval \* preds$se.fit)  
lwr <- preds$fit - (critval \* preds$se.fit)  
fit <- preds$fit  
  
  
#Simple plot of predicted values  
testData$predy = predict(model.final, testData, type = "response")  
### Plot  
plot(testData$Home ~ predicted,  
 pch = 16,  
 xlab="Predicted probability of 1 response",  
 ylab="Actual response")



#validate - confusion matrix  
confmatrix <- table(Actual\_Value=testData$Home, Predicted\_Value=predicted > 0.5)  
confmatrix

## Predicted\_Value  
## Actual\_Value FALSE TRUE  
## FALSE 102608 35900  
## TRUE 101029 253976

#sensitivity  
sensitivity(testData$Home, predicted)

## [1] 0.7154153

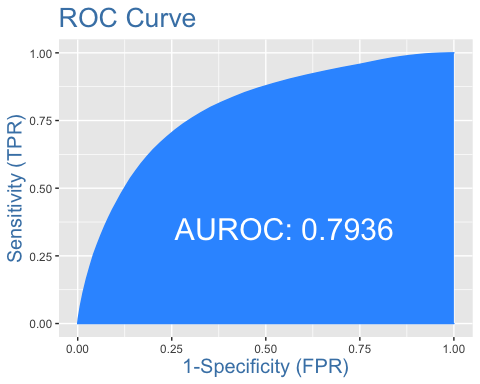
#specificity  
specificity(testData$Home, predicted)

## [1] 0.7408092

#accuracy  
(confmatrix[[1,1]] + confmatrix[[2,2]])/sum(confmatrix)

## [1] 0.7225423

plotROC(testData$Home, predicted)



#Check for overdispersion  
#One guideline is that if the ratio of the residual deviance to the residual degrees of freedom exceeds 1.5, then the model is overdispersed. Overdispersion indicates that the model doesn’t fit the data well: the explanatory variables may not well describe the dependent variable or the model may not be specified correctly for these data.  
summary(model.final)$deviance / summary(model.final)$df.residual

## [1] 1.101591

saveRDS(model.final, "home\_model.rds")

# Disposition to Other

Definition to “Disposition to Other” - Patients with a discharge status to any of the following:

print(unique(df[grepl("psychiatric hospital|short-term general hospital|Intermediate Care Facility|another type of institution|rehabilitation or long term|hospice care|inpatient rehab|Long Term Care Hospital|Skilled Nursing Facility", df$Discharge),]$Discharge))

## [1] "Discharged/transferred to a psychiatric hospital or psychiatric distin"  
## [2] "Discharged/Transferred to a short-term general hospital for inpatient"   
## [3] "Discharged/Transferred to an Intermediate Care Facility (ICF)"   
## [4] "Discharged/Transferred to another type of institution not defined else"  
## [5] "Discharged/Transferred to another type of rehabilitation or long term"   
## [6] "Discharged/Transferred to hospice care"   
## [7] "Discharged/Transferred to inpatient rehab or designated unit"   
## [8] "Discharged/Transferred to Long Term Care Hospital"   
## [9] "Discharged/Transferred to Skilled Nursing Facility"

#Check Class bias  
table(df$Other)

##   
## FALSE TRUE   
## 748759 391124

# Create Training Data  
input\_ones <- df[which(df$Other == TRUE), ] # all 1's  
input\_zeros <- df[which(df$Other == FALSE), ] # all 0's  
set.seed(100) # for repeatability of samples  
# Because there are more Falses than Trues, use 1s to set sample  
input\_ones\_training\_rows <- sample(1:nrow(input\_ones), 0.7\*nrow(input\_ones)) # 1's for training  
input\_zeros\_training\_rows <- sample(1:nrow(input\_zeros), 0.7\*nrow(input\_ones)) # 0's for training.  
training\_ones <- input\_ones[input\_ones\_training\_rows, ]   
training\_zeros <- input\_zeros[input\_zeros\_training\_rows, ]  
trainingData <- rbind(training\_ones, training\_zeros) # row bind the 1's and 0's   
  
# Create Test Data  
test\_ones <- input\_ones[-input\_ones\_training\_rows, ]  
test\_zeros <- input\_zeros[-input\_zeros\_training\_rows, ]  
testData <- rbind(test\_ones, test\_zeros) # row bind the 1's and 0's   
  
#Determining model with step procedure  
### Define full and null models and do step procedure  
model.full = glm(Other ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
 data=trainingData,   
 family=binomial(link="logit"))  
  
step(model.full,  
 scope = list(upper=model.full),  
 direction="both",  
 test="Chisq",  
 data=trainingData)

## Start: AIC=655808.1  
## Other ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
##   
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 655786 655808   
## - Alcohol 5 656250 656262 463 < 2.2e-16 \*\*\*  
## - GCSTOT 1 659036 659056 3250 < 2.2e-16 \*\*\*  
## - Gender 2 660089 660107 4303 < 2.2e-16 \*\*\*  
## - ISSAIS 1 664893 664913 9107 < 2.2e-16 \*\*\*  
## - Age 1 728410 728430 72624 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Call: glm(formula = Other ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
## family = binomial(link = "logit"), data = trainingData)  
##   
## Coefficients:  
## (Intercept)   
## -1.70492   
## Age   
## 0.04036   
## GenderMale   
## -0.41820   
## GenderNot Known/Not Recorded BIU 2   
## -0.21212   
## ISSAIS   
## 0.03683   
## AlcoholNo (not tested)   
## -0.06728   
## AlcoholNot Applicable BIU 1   
## -0.01455   
## AlcoholNot Known/Not Recorded BIU 2   
## 0.04402   
## AlcoholYes (confirmed by test [beyond legal limit])   
## -0.20554   
## AlcoholYes (confirmed by test [trace levels])   
## -0.06567   
## GCSTOT   
## -0.05526   
##   
## Degrees of Freedom: 547571 Total (i.e. Null); 547561 Residual  
## Null Deviance: 759100   
## Residual Deviance: 655800 AIC: 655800

#Final model  
model.final = glm(Other ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,  
 data=trainingData,  
 family = binomial(link="logit"),  
 na.action(na.omit)  
 )  
  
summary(model.final)

##   
## Call:  
## glm(formula = Other ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
## family = binomial(link = "logit"), data = trainingData, weights = na.action(na.omit))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.9982 -0.9837 -0.1339 0.9592 2.2155   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) -1.7049168 0.0197166  
## Age 0.0403624 0.0001613  
## GenderMale -0.4181962 0.0063707  
## GenderNot Known/Not Recorded BIU 2 -0.2121220 0.1617281  
## ISSAIS 0.0368258 0.0003964  
## AlcoholNo (not tested) -0.0672836 0.0071786  
## AlcoholNot Applicable BIU 1 -0.0145522 0.0175344  
## AlcoholNot Known/Not Recorded BIU 2 0.0440185 0.0168384  
## AlcoholYes (confirmed by test [beyond legal limit]) -0.2055444 0.0101975  
## AlcoholYes (confirmed by test [trace levels]) -0.0656706 0.0158210  
## GCSTOT -0.0552598 0.0009727  
## z value Pr(>|z|)   
## (Intercept) -86.471 < 2e-16 \*\*\*  
## Age 250.230 < 2e-16 \*\*\*  
## GenderMale -65.644 < 2e-16 \*\*\*  
## GenderNot Known/Not Recorded BIU 2 -1.312 0.18966   
## ISSAIS 92.896 < 2e-16 \*\*\*  
## AlcoholNo (not tested) -9.373 < 2e-16 \*\*\*  
## AlcoholNot Applicable BIU 1 -0.830 0.40658   
## AlcoholNot Known/Not Recorded BIU 2 2.614 0.00894 \*\*   
## AlcoholYes (confirmed by test [beyond legal limit]) -20.156 < 2e-16 \*\*\*  
## AlcoholYes (confirmed by test [trace levels]) -4.151 3.31e-05 \*\*\*  
## GCSTOT -56.811 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 759096 on 547571 degrees of freedom  
## Residual deviance: 655786 on 547561 degrees of freedom  
## AIC: 655808  
##   
## Number of Fisher Scoring iterations: 4

print(exp(cbind(OR = coef(model.final), confint(model.final))))

## OR 2.5 %  
## (Intercept) 0.1817875 0.1748941  
## Age 1.0411880 1.0408592  
## GenderMale 0.6582331 0.6500651  
## GenderNot Known/Not Recorded BIU 2 0.8088660 0.5891986  
## ISSAIS 1.0375123 1.0367071  
## AlcoholNo (not tested) 0.9349300 0.9218659  
## AlcoholNot Applicable BIU 1 0.9855532 0.9522740  
## AlcoholNot Known/Not Recorded BIU 2 1.0450017 1.0110845  
## AlcoholYes (confirmed by test [beyond legal limit]) 0.8142039 0.7980869  
## AlcoholYes (confirmed by test [trace levels]) 0.9364393 0.9078304  
## GCSTOT 0.9462392 0.9444366  
## 97.5 %  
## (Intercept) 0.1889474  
## Age 1.0415175  
## GenderMale 0.6665034  
## GenderNot Known/Not Recorded BIU 2 1.1116148  
## ISSAIS 1.0383193  
## AlcoholNo (not tested) 0.9481754  
## AlcoholNot Applicable BIU 1 1.0200300  
## AlcoholNot Known/Not Recorded BIU 2 1.0800738  
## AlcoholYes (confirmed by test [beyond legal limit]) 0.8306355  
## AlcoholYes (confirmed by test [trace levels]) 0.9659146  
## GCSTOT 0.9480445

#Analysis of variance for individual terms  
Anova(model.final, type="II", test="Wald")

## Analysis of Deviance Table (Type II tests)  
##   
## Response: Other  
## Df Chisq Pr(>Chisq)   
## Age 1 62614.93 < 2.2e-16 \*\*\*  
## Gender 2 4309.15 < 2.2e-16 \*\*\*  
## ISSAIS 1 8629.76 < 2.2e-16 \*\*\*  
## Alcohol 5 461.08 < 2.2e-16 \*\*\*  
## GCSTOT 1 3227.48 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Pseudo-R-squared  
nagelkerke(model.final)

## $Models  
##   
## Model: "glm, Other ~ Age + Gender + ISSAIS + Alcohol + GCSTOT, binomial(link = \"logit\"), trainingData, na.action(na.omit)"  
## Null: "glm, Other ~ 1, binomial(link = \"logit\"), trainingData, na.action(na.omit)"   
##   
## $Pseudo.R.squared.for.model.vs.null  
## Pseudo.R.squared  
## McFadden 0.136096  
## Cox and Snell (ML) 0.171939  
## Nagelkerke (Cragg and Uhler) 0.229253  
##   
## $Likelihood.ratio.test  
## Df.diff LogLik.diff Chisq p.value  
## -10 -51655 103310 0  
##   
## $Number.of.observations  
##   
## Model: 547572  
## Null: 547572  
##   
## $Messages  
## [1] "Note: For models fit with REML, these statistics are based on refitting with ML"  
##   
## $Warnings  
## [1] "None"

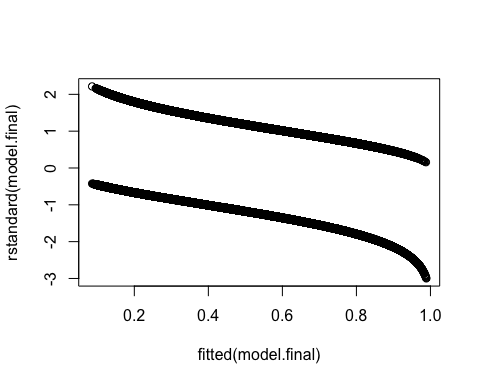
# Overall p-value for model  
### Define null models and compare to final model  
model.null = glm(Other ~ 1,  
 data=trainingData,  
 family = binomial(link="logit")  
 )  
  
anova(model.final,  
 model.null,  
 test="Chisq")

## Analysis of Deviance Table  
##   
## Model 1: Other ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
## Model 2: Other ~ 1  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 547561 655786   
## 2 547571 759096 -10 -103310 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

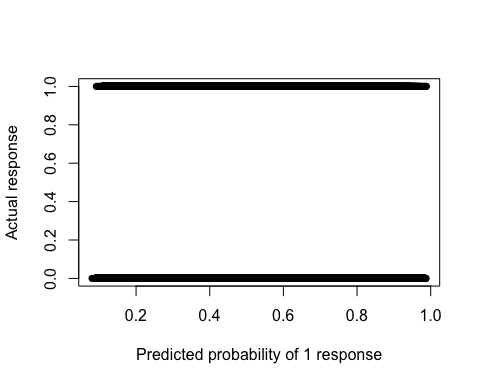
lrtest(model.final)

## Likelihood ratio test  
##   
## Model 1: Other ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
## Model 2: Other ~ 1  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 -327893   
## 2 1 -379548 -10 103310 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Plot of standardized residuals  
plot(fitted(model.final),  
 rstandard(model.final))



#prediction  
predicted <- predict(model.final, testData, type = "response")  
preds <- predict(model.final, newdata = testData, type = "response", se.fit = TRUE)  
critval <- 1.96 ## approx 95% CI  
upr <- preds$fit + (critval \* preds$se.fit)  
lwr <- preds$fit - (critval \* preds$se.fit)  
fit <- preds$fit  
  
#Simple plot of predicted values  
testData$predy = predict(model.final, testData, type = "response")  
### Plot  
plot(testData$Other ~ predicted,  
 pch = 16,  
 xlab="Predicted probability of 1 response",  
 ylab="Actual response")



#validate - confusion matrix  
confmatrix <- table(Actual\_Value=testData$Other, Predicted\_Value=predicted > 0.5)  
confmatrix

## Predicted\_Value  
## Actual\_Value FALSE TRUE  
## FALSE 319771 155202  
## TRUE 33506 83832

#sensitivity  
sensitivity(testData$Other, predicted)

## [1] 0.7144489

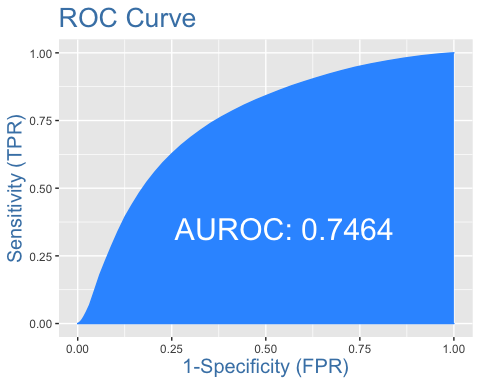
#specificity  
specificity(testData$Other, predicted)

## [1] 0.6732404

#accuracy  
(confmatrix[[1,1]] + confmatrix[[2,2]])/sum(confmatrix)

## [1] 0.6814039

plotROC(testData$Other, predicted)



#Check for overdispersion  
#One guideline is that if the ratio of the residual deviance to the residual degrees of freedom exceeds 1.5, then the model is overdispersed. Overdispersion indicates that the model doesn’t fit the data well: the explanatory variables may not well describe the dependent variable or the model may not be specified correctly for these data.  
summary(model.final)$deviance / summary(model.final)$df.residual

## [1] 1.197649

saveRDS(model.final, "other\_model.rds")

# Tracheostomy

#Check Class bias  
table(df$Tracheostomy)

##   
## FALSE TRUE   
## 1111839 28044

# Create Training Data  
input\_ones <- df[which(df$Tracheostomy == TRUE), ] # all 1's  
input\_zeros <- df[which(df$Tracheostomy == FALSE), ] # all 0's  
set.seed(100) # for repeatability of samples  
# Because there are more Falses than Trues, use 1s to set sample  
input\_ones\_training\_rows <- sample(1:nrow(input\_ones), 0.7\*nrow(input\_ones)) # 1's for training  
input\_zeros\_training\_rows <- sample(1:nrow(input\_zeros), 0.7\*nrow(input\_ones)) # 0's for training.  
training\_ones <- input\_ones[input\_ones\_training\_rows, ]   
training\_zeros <- input\_zeros[input\_zeros\_training\_rows, ]  
trainingData <- rbind(training\_ones, training\_zeros) # row bind the 1's and 0's   
  
# Create Test Data  
test\_ones <- input\_ones[-input\_ones\_training\_rows, ]  
test\_zeros <- input\_zeros[-input\_zeros\_training\_rows, ]  
testData <- rbind(test\_ones, test\_zeros) # row bind the 1's and 0's   
  
#Determining model with step procedure  
### Define full and null models and do step procedure  
model.full = glm(Tracheostomy ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,   
 data=trainingData,   
 family=binomial(link="logit"))  
  
step(model.full,  
 scope = list(upper=model.full),  
 direction="both",  
 test="Chisq",  
 data=trainingData)

## Start: AIC=38288.45  
## Tracheostomy ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
##   
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 38266 38288   
## - Age 1 38437 38457 170.1 < 2.2e-16 \*\*\*  
## - Gender 2 38453 38471 186.2 < 2.2e-16 \*\*\*  
## - Alcohol 5 38490 38502 223.6 < 2.2e-16 \*\*\*  
## - ISSAIS 1 42583 42603 4316.7 < 2.2e-16 \*\*\*  
## - GCSTOT 1 42744 42764 4477.1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Call: glm(formula = Tracheostomy ~ Age + Gender + ISSAIS + Alcohol +   
## GCSTOT, family = binomial(link = "logit"), data = trainingData)  
##   
## Coefficients:  
## (Intercept)   
## -0.063799   
## Age   
## 0.008507   
## GenderMale   
## 0.388325   
## GenderNot Known/Not Recorded BIU 2   
## 1.095429   
## ISSAIS   
## 0.086883   
## AlcoholNo (not tested)   
## -0.366181   
## AlcoholNot Applicable BIU 1   
## -0.434495   
## AlcoholNot Known/Not Recorded BIU 2   
## 0.009215   
## AlcoholYes (confirmed by test [beyond legal limit])   
## 0.040620   
## AlcoholYes (confirmed by test [trace levels])   
## 0.173255   
## GCSTOT   
## -0.188296   
##   
## Degrees of Freedom: 39259 Total (i.e. Null); 39249 Residual  
## Null Deviance: 54430   
## Residual Deviance: 38270 AIC: 38290

#Final model  
model.final = glm(Tracheostomy ~ Age + Gender + ISSAIS + Alcohol + GCSTOT,  
 data=trainingData,  
 family = binomial(link="logit"),  
 na.action(na.omit)  
 )  
summary(model.final)

##   
## Call:  
## glm(formula = Tracheostomy ~ Age + Gender + ISSAIS + Alcohol +   
## GCSTOT, family = binomial(link = "logit"), data = trainingData,   
## weights = na.action(na.omit))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.7325 -0.7275 -0.1483 0.7346 2.2357   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) -0.0637988 0.0671063  
## Age 0.0085075 0.0006565  
## GenderMale 0.3883252 0.0287120  
## GenderNot Known/Not Recorded BIU 2 1.0954293 0.6110853  
## ISSAIS 0.0868831 0.0014791  
## AlcoholNo (not tested) -0.3661813 0.0301579  
## AlcoholNot Applicable BIU 1 -0.4344946 0.0838707  
## AlcoholNot Known/Not Recorded BIU 2 0.0092151 0.0712577  
## AlcoholYes (confirmed by test [beyond legal limit]) 0.0406200 0.0379868  
## AlcoholYes (confirmed by test [trace levels]) 0.1732555 0.0582042  
## GCSTOT -0.1882961 0.0030545  
## z value Pr(>|z|)   
## (Intercept) -0.951 0.34175   
## Age 12.959 < 2e-16 \*\*\*  
## GenderMale 13.525 < 2e-16 \*\*\*  
## GenderNot Known/Not Recorded BIU 2 1.793 0.07304 .   
## ISSAIS 58.742 < 2e-16 \*\*\*  
## AlcoholNo (not tested) -12.142 < 2e-16 \*\*\*  
## AlcoholNot Applicable BIU 1 -5.181 2.21e-07 \*\*\*  
## AlcoholNot Known/Not Recorded BIU 2 0.129 0.89710   
## AlcoholYes (confirmed by test [beyond legal limit]) 1.069 0.28493   
## AlcoholYes (confirmed by test [trace levels]) 2.977 0.00291 \*\*   
## GCSTOT -61.645 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 54426 on 39259 degrees of freedom  
## Residual deviance: 38266 on 39249 degrees of freedom  
## AIC: 38288  
##   
## Number of Fisher Scoring iterations: 5

print(exp(cbind(OR = coef(model.final), confint(model.final))))

## OR 2.5 %  
## (Intercept) 0.9381938 0.8226091  
## Age 1.0085438 1.0072483  
## GenderMale 1.4745092 1.3938803  
## GenderNot Known/Not Recorded BIU 2 2.9904662 0.8871714  
## ISSAIS 1.0907692 1.0876234  
## AlcoholNo (not tested) 0.6933771 0.6535747  
## AlcoholNot Applicable BIU 1 0.6475918 0.5488271  
## AlcoholNot Known/Not Recorded BIU 2 1.0092577 0.8773710  
## AlcoholYes (confirmed by test [beyond legal limit]) 1.0414563 0.9667406  
## AlcoholYes (confirmed by test [trace levels]) 1.1891699 1.0610316  
## GCSTOT 0.8283694 0.8234067  
## 97.5 %  
## (Intercept) 1.0701373  
## Age 1.0098438  
## GenderMale 1.5599340  
## GenderNot Known/Not Recorded BIU 2 10.1580735  
## ISSAIS 1.0939477  
## AlcoholNo (not tested) 0.7355927  
## AlcoholNot Applicable BIU 1 0.7625096  
## AlcoholNot Known/Not Recorded BIU 2 1.1601312  
## AlcoholYes (confirmed by test [beyond legal limit]) 1.1219693  
## AlcoholYes (confirmed by test [trace levels]) 1.3329786  
## GCSTOT 0.8333255

#Analysis of variance for individual terms  
Anova(model.final, type="II", test="Wald")

## Analysis of Deviance Table (Type II tests)  
##   
## Response: Tracheostomy  
## Df Chisq Pr(>Chisq)   
## Age 1 167.93 < 2.2e-16 \*\*\*  
## Gender 2 184.77 < 2.2e-16 \*\*\*  
## ISSAIS 1 3450.67 < 2.2e-16 \*\*\*  
## Alcohol 5 223.42 < 2.2e-16 \*\*\*  
## GCSTOT 1 3800.12 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Pseudo-R-squared  
nagelkerke(model.final)

## $Models  
##   
## Model: "glm, Tracheostomy ~ Age + Gender + ISSAIS + Alcohol + GCSTOT, binomial(link = \"logit\"), trainingData, na.action(na.omit)"  
## Null: "glm, Tracheostomy ~ 1, binomial(link = \"logit\"), trainingData, na.action(na.omit)"   
##   
## $Pseudo.R.squared.for.model.vs.null  
## Pseudo.R.squared  
## McFadden 0.296908  
## Cox and Snell (ML) 0.337412  
## Nagelkerke (Cragg and Uhler) 0.449882  
##   
## $Likelihood.ratio.test  
## Df.diff LogLik.diff Chisq p.value  
## -10 -8079.7 16159 0  
##   
## $Number.of.observations  
##   
## Model: 39260  
## Null: 39260  
##   
## $Messages  
## [1] "Note: For models fit with REML, these statistics are based on refitting with ML"  
##   
## $Warnings  
## [1] "None"

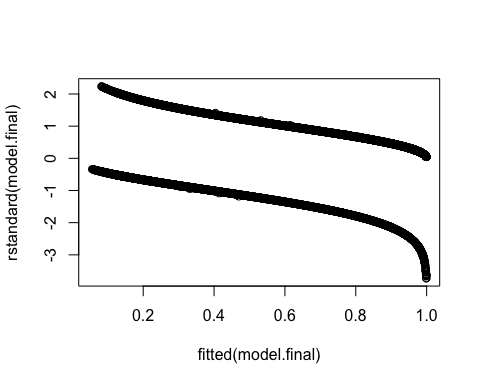
# Overall p-value for model  
### Define null models and compare to final model  
model.null = glm(Tracheostomy ~ 1,  
 data=trainingData,  
 family = binomial(link="logit")  
 )  
  
anova(model.final,  
 model.null,  
 test="Chisq")

## Analysis of Deviance Table  
##   
## Model 1: Tracheostomy ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
## Model 2: Tracheostomy ~ 1  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 39249 38266   
## 2 39259 54426 -10 -16160 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

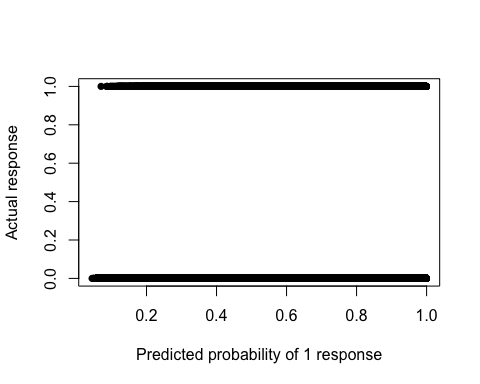
lrtest(model.final)

## Likelihood ratio test  
##   
## Model 1: Tracheostomy ~ Age + Gender + ISSAIS + Alcohol + GCSTOT  
## Model 2: Tracheostomy ~ 1  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 -19133   
## 2 1 -27213 -10 16159 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Plot of standardized residuals  
plot(fitted(model.final),  
 rstandard(model.final))



#prediction  
predicted <- predict(model.final, testData, type = "response")  
preds <- predict(model.final, newdata = testData, type = "response", se.fit = TRUE)  
critval <- 1.96 ## approx 95% CI  
upr <- preds$fit + (critval \* preds$se.fit)  
lwr <- preds$fit - (critval \* preds$se.fit)  
fit <- preds$fit  
  
#Simple plot of predicted values  
testData$predy = predict(model.final, testData, type = "response")  
### Plot  
plot(testData$Tracheostomy ~ predicted,  
 pch = 16,  
 xlab="Predicted probability of 1 response",  
 ylab="Actual response")



#validate - confusion matrix  
confmatrix <- table(Actual\_Value=testData$Tracheostomy, Predicted\_Value=predicted > 0.5)  
confmatrix

## Predicted\_Value  
## Actual\_Value FALSE TRUE  
## FALSE 916390 175819  
## TRUE 2225 6189

#sensitivity  
sensitivity(testData$Tracheostomy, predicted)

## [1] 0.7355598

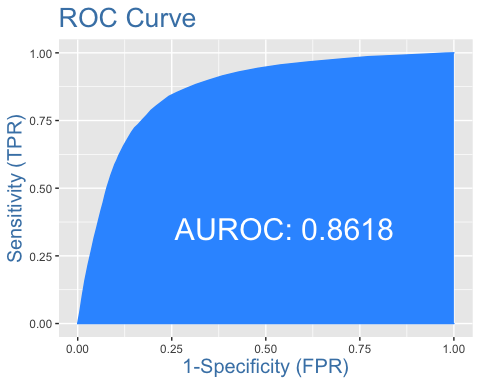
#specificity  
specificity(testData$Tracheostomy, predicted)

## [1] 0.8390244

#accuracy  
(confmatrix[[1,1]] + confmatrix[[2,2]])/sum(confmatrix)

## [1] 0.8382334

plotROC(testData$Tracheostomy, predicted)



#Check for overdispersion  
#One guideline is that if the ratio of the residual deviance to the residual degrees of freedom exceeds 1.5, then the model is overdispersed. Overdispersion indicates that the model doesn’t fit the data well: the explanatory variables may not well describe the dependent variable or the model may not be specified correctly for these data.  
summary(model.final)$deviance / summary(model.final)$df.residual

## [1] 0.9749663

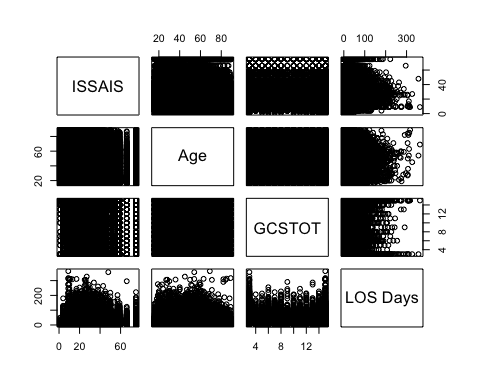
saveRDS(model.final, "tracheostomy\_model.rds")

# Length of Stay Prediction

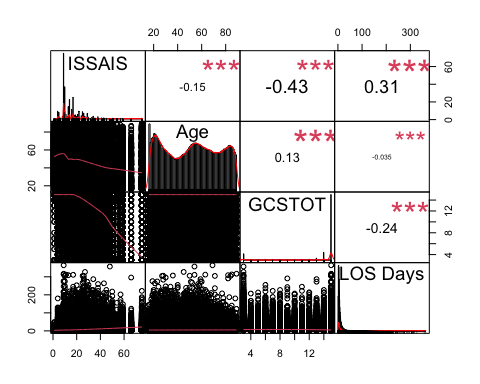
data.los =  
 select(df,  
 ISSAIS,  
 Alcohol,  
 Gender,  
 Age,  
 GCSTOT,  
 `LOS Days`)  
  
library(psych)  
corr.test(data.los[,-2:-3],  
 use = "pairwise",  
 method="pearson",  
 adjust="none", # Can adjust p-values; see ?p.adjust for options  
 alpha=.05)

## Call:corr.test(x = data.los[, -2:-3], use = "pairwise", method = "pearson",   
## adjust = "none", alpha = 0.05)  
## Correlation matrix   
## ISSAIS Age GCSTOT LOS Days  
## ISSAIS 1.00 -0.15 -0.43 0.31  
## Age -0.15 1.00 0.13 -0.03  
## GCSTOT -0.43 0.13 1.00 -0.24  
## LOS Days 0.31 -0.03 -0.24 1.00  
## Sample Size   
## [1] 1139883  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## ISSAIS Age GCSTOT LOS Days  
## ISSAIS 0 0 0 0  
## Age 0 0 0 0  
## GCSTOT 0 0 0 0  
## LOS Days 0 0 0 0  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

pairs(data=data.los,  
 ~ ISSAIS + Age + GCSTOT + `LOS Days`)



library(PerformanceAnalytics)  
chart.Correlation(data.los[,-2:-3],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)



#Training and Test Set  
set.seed(100)  
row.number <- sample(1:nrow(data.los), 0.8\*nrow(data.los))  
trainingData = data.los[row.number,]  
testData = data.los[-row.number,]  
dim(trainingData)

## [1] 911906 6

dim(testData)

## [1] 227977 6

# Multiple regression  
# Stepwise procedure  
model.null = lm(`LOS Days` ~ 1,  
 data=trainingData)  
model.full = lm(`LOS Days` ~ ISSAIS + Alcohol + Gender + Age + GCSTOT,  
 data=trainingData)  
step(model.null,  
 scope = list(upper=model.full),  
 direction="both",  
 data=trainingData)

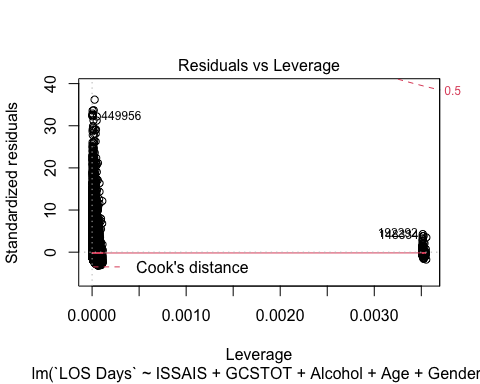
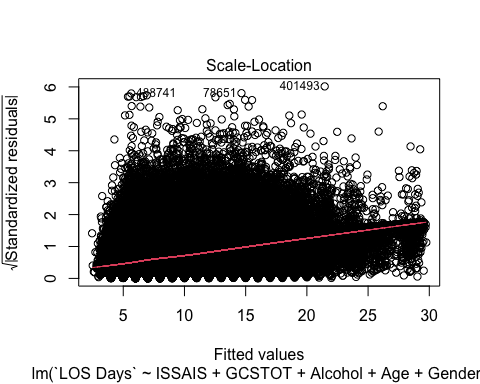
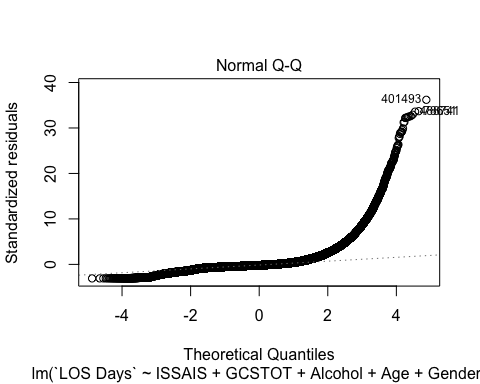
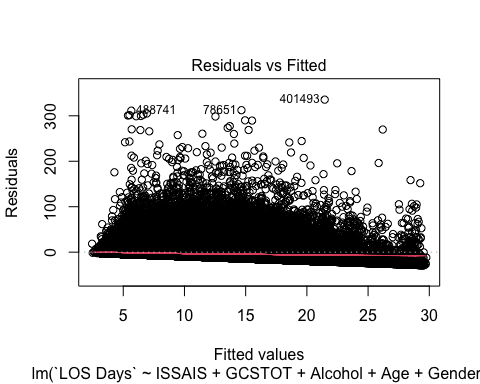
## Start: AIC=4167268  
## `LOS Days` ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + ISSAIS 1 8127599 79897572 4078927  
## + GCSTOT 1 5228434 82796737 4111430  
## + Alcohol 5 823330 87201841 4158708  
## + Gender 2 218287 87806885 4165007  
## + Age 1 104051 87921120 4166191  
## <none> 88025171 4167268  
##   
## Step: AIC=4078927  
## `LOS Days` ~ ISSAIS  
##   
## Df Sum of Sq RSS AIC  
## + GCSTOT 1 1354511 78543061 4063336  
## + Alcohol 5 138492 79759080 4077355  
## + Gender 2 38515 79859057 4078491  
## + Age 1 9330 79888242 4078822  
## <none> 79897572 4078927  
## - ISSAIS 1 8127599 88025171 4167268  
##   
## Step: AIC=4063336  
## `LOS Days` ~ ISSAIS + GCSTOT  
##   
## Df Sum of Sq RSS AIC  
## + Alcohol 5 78439 78464622 4062435  
## + Age 1 35684 78507377 4062924  
## + Gender 2 22342 78520719 4063081  
## <none> 78543061 4063336  
## - GCSTOT 1 1354511 79897572 4078927  
## - ISSAIS 1 4253676 82796737 4111430  
##   
## Step: AIC=4062435  
## `LOS Days` ~ ISSAIS + GCSTOT + Alcohol  
##   
## Df Sum of Sq RSS AIC  
## + Age 1 70595 78394028 4061616  
## + Gender 2 13309 78451314 4062285  
## <none> 78464622 4062435  
## - Alcohol 5 78439 78543061 4063336  
## - GCSTOT 1 1294457 79759080 4077355  
## - ISSAIS 1 4017787 82482409 4107971  
##   
## Step: AIC=4061616  
## `LOS Days` ~ ISSAIS + GCSTOT + Alcohol + Age  
##   
## Df Sum of Sq RSS AIC  
## + Gender 2 31519 78362509 4061254  
## <none> 78394028 4061616  
## - Age 1 70595 78464622 4062435  
## - Alcohol 5 113350 78507377 4062924  
## - GCSTOT 1 1320537 79714564 4076847  
## - ISSAIS 1 4071307 82465335 4107784  
##   
## Step: AIC=4061254  
## `LOS Days` ~ ISSAIS + GCSTOT + Alcohol + Age + Gender  
##   
## Df Sum of Sq RSS AIC  
## <none> 78362509 4061254  
## - Gender 2 31519 78394028 4061616  
## - Age 1 88805 78451314 4062285  
## - Alcohol 5 103042 78465551 4062442  
## - GCSTOT 1 1316411 79678920 4076444  
## - ISSAIS 1 4039965 82402474 4107093

##   
## Call:  
## lm(formula = `LOS Days` ~ ISSAIS + GCSTOT + Alcohol + Age + Gender,   
## data = trainingData)  
##   
## Coefficients:  
## (Intercept)   
## 8.49078   
## ISSAIS   
## 0.27306   
## GCSTOT   
## -0.39746   
## AlcoholNo (not tested)   
## -0.71762   
## AlcoholNot Applicable BIU 1   
## -0.65970   
## AlcoholNot Known/Not Recorded BIU 2   
## -0.18811   
## AlcoholYes (confirmed by test [beyond legal limit])   
## -0.19984   
## AlcoholYes (confirmed by test [trace levels])   
## 0.33849   
## Age   
## 0.01561   
## GenderMale   
## 0.40680   
## GenderNot Known/Not Recorded BIU 2   
## 0.34768

#Define final model  
model.final = lm(`LOS Days` ~ ISSAIS + GCSTOT + Alcohol + Age + Gender,  
 data=trainingData)  
summary(model.final)

##   
## Call:  
## lm(formula = `LOS Days` ~ ISSAIS + GCSTOT + Alcohol + Age + Gender,   
## data = trainingData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -28.57 -3.96 -1.80 1.34 335.54   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 8.4907806 0.0644145  
## ISSAIS 0.2730576 0.0012594  
## GCSTOT -0.3974649 0.0032113  
## AlcoholNo (not tested) -0.7176152 0.0232751  
## AlcoholNot Applicable BIU 1 -0.6596961 0.0582941  
## AlcoholNot Known/Not Recorded BIU 2 -0.1881073 0.0552572  
## AlcoholYes (confirmed by test [beyond legal limit]) -0.1998403 0.0321202  
## AlcoholYes (confirmed by test [trace levels]) 0.3384873 0.0499574  
## Age 0.0156144 0.0004857  
## GenderMale 0.4068024 0.0212420  
## GenderNot Known/Not Recorded BIU 2 0.3476819 0.5493656  
## t value Pr(>|t|)   
## (Intercept) 131.815 < 2e-16 \*\*\*  
## ISSAIS 216.824 < 2e-16 \*\*\*  
## GCSTOT -123.770 < 2e-16 \*\*\*  
## AlcoholNo (not tested) -30.832 < 2e-16 \*\*\*  
## AlcoholNot Applicable BIU 1 -11.317 < 2e-16 \*\*\*  
## AlcoholNot Known/Not Recorded BIU 2 -3.404 0.000664 \*\*\*  
## AlcoholYes (confirmed by test [beyond legal limit]) -6.222 4.92e-10 \*\*\*  
## AlcoholYes (confirmed by test [trace levels]) 6.776 1.24e-11 \*\*\*  
## Age 32.147 < 2e-16 \*\*\*  
## GenderMale 19.151 < 2e-16 \*\*\*  
## GenderNot Known/Not Recorded BIU 2 0.633 0.526813   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.27 on 911895 degrees of freedom  
## Multiple R-squared: 0.1098, Adjusted R-squared: 0.1098   
## F-statistic: 1.124e+04 on 10 and 911895 DF, p-value: < 2.2e-16

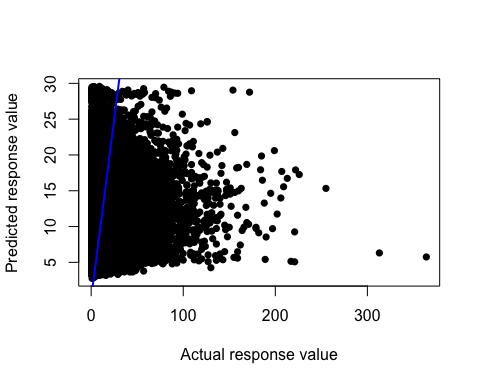
plot(model.final)



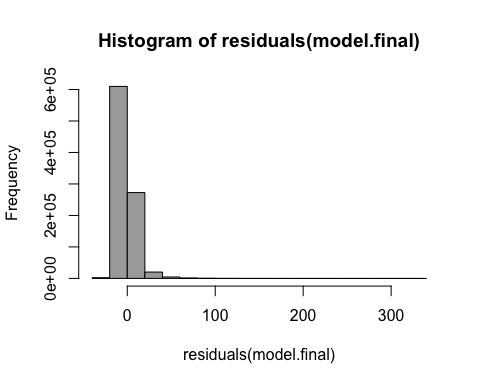
#Analysis of variance for individual terms  
Anova(model.final, Type="II")

## Anova Table (Type II tests)  
##   
## Response: LOS Days  
## Sum Sq Df F value Pr(>F)   
## ISSAIS 4039965 1 47012.58 < 2.2e-16 \*\*\*  
## GCSTOT 1316411 1 15318.91 < 2.2e-16 \*\*\*  
## Alcohol 103042 5 239.82 < 2.2e-16 \*\*\*  
## Age 88805 1 1033.41 < 2.2e-16 \*\*\*  
## Gender 31519 2 183.39 < 2.2e-16 \*\*\*  
## Residuals 78362509 911895   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

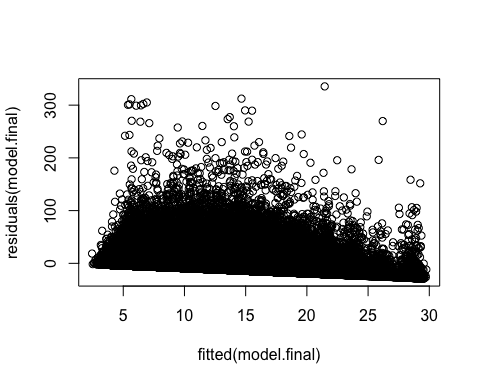
#Simple plot of predicted values with 1-to-1 line  
testData$predy = predict(model.final,testData, se.fit = TRUE)$fit  
plot(predy ~ `LOS Days`,  
 data=testData,  
 pch = 16,  
 xlab="Actual response value",  
 ylab="Predicted response value")  
abline(0,1, col="blue", lwd=2)



#Checking assumptions of the model  
hist(residuals(model.final),  
 col="darkgray")



#A histogram of residuals from a linear model. The distribution of these residuals should be approximately normal.  
plot(fitted(model.final),  
 residuals(model.final)  
 )



#A plot of residuals vs. predicted values. The residuals should be unbiased and homoscedastic. For an illustration of these properties, see this diagram by Steve Jost at DePaul University: condor.depaul.edu/sjost/it223/documents/resid-plots.gif.  
saveRDS(model.final, "los\_model.rds")