```
setwd('C:/R')
load('NatalRiskData.Rdata')
head(sdata)

train <- sdata[sdata$ORIGRANDGROUP<=5,]
test <- sdata[sdata$ORIGRANDGROUP>5,]

names(train)
names(test)

rm(sdata)

complications <- c("ULD_MECO","ULD_PRECIP","ULD_BREECH")

riskfactors <- c("URF_DIAB", "URF_CHYPER", "URF_PHYPER",
                 "URF_ECLAM")

y <- factor(train$atRisk)

x <- c("PWGT",
       "UPREVIS",
       "CIG_REC",
       "GESTREC3",
       "DPLURAL",
       complications,
       riskfactors)
fmla <- paste("atRisk", paste(x, collapse="+"), sep="~")

print(fmla)
```

```
log_reg <- glm(fmla, data = train, family = "binomial")
summary(log_reg)

train$pred <- predict(log_reg, newdata = train, type = "response")
head(train$pred)

table(y)
sum(train$pred)

library(ggplot2)

ggplot(train, aes(x=pred, color=atRisk, linetype=atRisk)) +
  geom_density()

test$pred <- predict(log_reg, newdata=test, type="response")
head(test)
summary(test$pred)

ggplot(test, aes(x=pred, color=atRisk, linetype=atRisk)) +
  geom_density()

confusion.test <- table(pred = test$pred>0.02, target = test$atRisk)
confusion.test

# Confusion Matrix
#       target
# pred  FALSE TRUE
# FALSE  9487   93
```

```
# TRUE   2405  116


#Then calculate, accuracy, precision and recall
accuracy <- (confusion.test[2,2] + confusion.test[1,1])/sum(confusion.test[,])
accuracy    # 0.7935708
precision <- confusion.test[2,2] / sum(confusion.test[2,])
precision   # 0.04601349
recall <- confusion.test[2,2] / sum(confusion.test[,2])
recall      # 0.5550239
fit1 <- lm( ssc ~ age + location + ethnicity + coder + som1 + som2 + som3 +
            som4 + som5 + som10 + som11 + som12 + som13 + som14)
summary(fit1)
detach(trainSet)
rm(trainSet)
testSet$ssc_pred <- predict(fit1, newdata = testSet)
rm(fit1)
library(ggplot2)
ggplot(data = testSet, aes(x = ssc_pred, y = ssc)) +
  geom_point(color = "red") +
  geom_line(aes(x = ssc, y = ssc), color = "blue")
```