```r
library("caret")
library("IDPmisc")
library("magrittr")
library("dplyr")
library("tidyr")
library("lmtest")
library("popbio")
library("e1071")

minerals$GoldYN = NA
minerals$GoldYN[minerals$Gold == '1'] = 1
minerals$GoldYN[minerals$Gold == '0'] = 0

MineralsGold = NaRV.omit(minerals)


mylogit = glm(GoldYN~Antimony, data=MineralsGold, family="binomial")

probabilities <- predict(mylogit, type = "response")
MineralsGold$Predicted = ifelse(probabilities > .05, "pos", "neg")

MineralsGold$PredictedR = NA
MineralsGold$PredictedR[MineralsGold$Predicted == 'pos'] = 1
MineralsGold$PredictedR[MineralsGold$Predicted == 'neg'] = 0

MineralsGold$PredictedR = as.factor(MineralsGold$PredictedR)
MineralsGold$GoldYN = as.factor(MineralsGold$GoldYN)

#Sample Size
conf_matrix = caret::confusionMatrix(MineralsGold$PredictedR, MineralsGold$GoldYN)
conf_matrix
# Don't meet the assumption for sample size

# Logit Linearity
MineralsGold1 = MineralsGold %>% dplyr::select_if(is.numeric)
predictors = colnames(MineralsGold1)

MineralsGold2 = MineralsGold1 %>%
  mutate(logit=log(probabilities/(1-probabilities))) %>%
  gather(key= "predictors", value="predictor.value", -logit)

ggplot(MineralsGold2, aes(logit, predictor.value)) +
  geom_point(size=.5, alpha=.5) +
  geom_smooth(method = "loess") +
```

```r
  theme_bw() +
  facet_wrap(~predictors, scales="free_y")
# it is linear

# Multicollinearity = only test if you have multiple IVs

# Independent Errors
plot(mylogit$residuals)

dwtest(mylogit, alternative="two.sided")
# greater than .05, meeting this assumption

# outliers
infl <- influence.measures(mylogit)
summary(infl)
# not many

summary(mylogit)
#significant for the output

logi.hist.plot(MineralsGold$Antimony, MineralsGold$GoldYN, boxp = FALSE, type = "hist")

#low probability
```