```
#title: Week 10 Logistic regression

#author: Vasanthakumar Kalaikkovan

#Date: 05/22/2021


getwd()


setwd("E://Repos/StatisticsR/DSC520-Statistics/week10")


install.packages("farff")


library(farff)


surgery_df <- readARFF("ThoraricSurgery.arff")


head(surgery_df)
```

Fit a binary logistic regression model to the data set that predicts whether the patient survived for one year (the Risk1Y variable) after the surgery. Use the glm() function to perform the logistic regression. See Generalized Linear Models for an example. Include a summary using the summary() function in your results.

```
survived_model <- glm(Risk1Yr ~
PRE4+PRE5+AGE+DGN+PRE6+PRE7+PRE8+PRE9+PRE10+PRE11+PRE14+

PRE17+PRE19+PRE25+PRE30+PRE32,data=surgery_df,family=binomial())
```

According to the summary, which variables had the greatest effect on the survival rate?

```
summary(survived_model)
```

Answer - For p <0.05, We are having DGNDGN5,DGNDGN8, PRE9F, PRE14OC14, PRE17F, and PRE30F, half of which have negative coefficients.

To compute the accuracy of your model, use the dataset to predict the outcome variable. The percent of correct predictions is the accuracy of your model. What is the accuracy of your model?

```
psuedo_R2 <- 1-(survived_model$deviance/survived_model$null.deviance)

psuedo_R2
```

Answer - Pseudo R-squared even with all variables is low at 0.14, Residual deviance, 341.2 is not much smaller than null deviance. The accuracy of the model is not very good, and most variables have high p-values, including the intercept. Lot of models need to reinvestigate, multicollinearity not addressed. AIC of 391.2 can be used to compare with other models.