---  
title: "ASSIGNMENT 7.1\_ThoraricSurgery"  
author: "Bhushan Suryawanshi"  
date: '2020-07-13'  
  
---

For this problem, you will be working with the thoracic surgery data set from the University of California Irvine machine learning repository. This dataset contains information on life expectancy in lung cancer patients after surgery.

The underlying thoracic surgery data is in ARFF format. This is a text-based format with information on each of the attributes. You can load this data using a package such as foreign or by cutting and pasting the data section into a CSV file.

Assignment Instructions:

Include all of your answers in a R Markdown report. Here is an example R Markdown report that you can use as a guide.

library("foreign")  
thoracic\_surgery\_df <- read.arff("ThoraricSurgery.arff")  
head(thoracic\_surgery\_df)

## DGN PRE4 PRE5 PRE6 PRE7 PRE8 PRE9 PRE10 PRE11 PRE14 PRE17 PRE19 PRE25 PRE30  
## 1 DGN2 2.88 2.16 PRZ1 F F F T T OC14 F F F T  
## 2 DGN3 3.40 1.88 PRZ0 F F F F F OC12 F F F T  
## 3 DGN3 2.76 2.08 PRZ1 F F F T F OC11 F F F T  
## 4 DGN3 3.68 3.04 PRZ0 F F F F F OC11 F F F F  
## 5 DGN3 2.44 0.96 PRZ2 F T F T T OC11 F F F T  
## 6 DGN3 2.48 1.88 PRZ1 F F F T F OC11 F F F F  
## PRE32 AGE Risk1Yr  
## 1 F 60 F  
## 2 F 51 F  
## 3 F 59 F  
## 4 F 54 F  
## 5 F 73 T  
## 6 F 51 F

1. Fit a binary logistic regression model to the data set that predicts whether or not 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.
2. According to the summary, which variables had the greatest effect on the survival rate?
3. 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?