

```

# Week 3, The Framingham Heart Study

# Video 3

# Read in the dataset
framingham = read.csv("framingham.csv")

# Look at structure
str(framingham)

# Load the library caTools
library(caTools)

# Randomly split the data into training and testing sets
set.seed(1000)
split = sample.split(framingham$TenYearCHD, SplitRatio = 0.65)

# Split up the data using subset
train = subset(framingham, split==TRUE)
test = subset(framingham, split==FALSE)

# Logistic Regression Model
framinghamLog = glm(TenYearCHD ~ ., data = train, family=binomial)
summary(framinghamLog)

# Predictions on the test set
predictTest = predict(framinghamLog, type="response", newdata=test)

# Confusion matrix with threshold of 0.5
table(test$TenYearCHD, predictTest > 0.5)

# Accuracy
(1069+11)/(1069+6+187+11)

# Baseline accuracy
(1069+6)/(1069+6+187+11)

# Test set AUC
library(ROCR)
ROCRpred = prediction(predictTest, test$TenYearCHD)
as.numeric(performance(ROCRpred, "auc")@y.values)

```