```
> #Multiple Logistic Regression in R
> library(epiDisplay)
> library(pROC)
> library(caret)
> library(faraway)
> library(StepReg)
>
> if (FALSE)
+ {"
+ The National Institute of Diabetes and Digestive and Kidney Diseases
+ conducted a study on 768 adult female Pima Indians living near Phoenix.
+ The pima dataset available in R contains the following variables
+
+ test - results of a test to determine if the female patient shows signs of diabetes (coded 0 if
negative, 1 if positive)
+ age - Age (years)
+ bmi - Body mass index (weight in kg/(height in metres squared))
+ diastolic - Diastolic blood pressure (mm Hg)
+ diabetes - Diabetes pedigree function
+ glucose - Plasma glucose concentration at 2 hours in an oral glucose tolerance test
+ insulin - 2-Hour serum insulin (mu U/ml)
+ pregnant - Number of times pregnant
+ triceps - Triceps skin fold thickness (mm)
+ "}
>
> library(faraway)
> library(StepReg)
>
> logistic <- glm(test ~ age + bmi + diastolic + diabetes + glucose + insulin + pregnant +
triceps,family=binomial(logit),data=pima)
> summary(logistic)

Call:
glm(formula = test ~ age + bmi + diastolic + diabetes + glucose +
    insulin + pregnant + triceps, family = binomial(logit), data = pima)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5566  -0.7274  -0.4159   0.7267   2.9297

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.4046964  0.7166359 -11.728  < 2e-16 ***
age          0.0148690  0.0093348   1.593 0.111192
bmi          0.0897010  0.0150876   5.945 2.76e-09 ***
diastolic   -0.0132955  0.0052336  -2.540 0.011072 *
diabetes     0.9451797  0.2991475   3.160 0.001580 **
glucose      0.0351637  0.0037087   9.481  < 2e-16 ***
insulin     -0.0011917  0.0009012  -1.322 0.186065
pregnant     0.1231823  0.0320776   3.840 0.000123 ***
triceps      0.0006190  0.0068994   0.090 0.928515
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 993.48  on 767  degrees of freedom
Residual deviance: 723.45  on 759  degrees of freedom
AIC: 741.45

Number of Fisher Scoring iterations: 5

> anova(logistic)
Analysis of Deviance Table
```

```
Model: binomial, link: logit

Response: test

Terms added sequentially (first to last)


         Df Deviance Resid. Df Resid. Dev
NULL                       767      993.48
age        1   42.764      766      950.72
bmi        1   75.045      765      875.68
diastolic  1    3.782      764      871.89
diabetes   1   13.529      763      858.36
glucose    1  117.098      762      741.27
insulin    1    2.574      761      738.69
pregnant   1   15.239      760      723.45
triceps    1    0.008      759      723.45
>
> windows(7,7)
> #save graph(s) in pdf
>
pdf(file="C:/Users/jmard/OneDrive/Desktop/RegressionMethodsSpring2020/Homework/HW09_Figures.pdf")
>
> predicted.test <- ifelse(logistic$fitted.values > 0.5,1,0)
>
> predicted.test <- as.factor(predicted.test) #need to be sure this variable is a factor
> pima$test <- as.factor(pima$test)  #need to be sure this variable is a factor
>
> table(pima$test)

  0   1
500 268
> table(predicted.test)
predicted.test
  0   1
557 211
>
> #These results are slightly different than results in HW
> #Also, the confusion matrix from this output considers '0' as the event
> confusionMatrix(predicted.test,pima$test,positive='1')  #found in the caret library
Confusion Matrix and Statistics
```

For HW 9

|            | Reference |     |
|------------|-----------|-----|
| Prediction | 0         | 1   |
| 0          | 446       | 117 |
| 1          | 54        | 151 |

| | |
|---|---|
| Accuracy | (446+151)/768 = 0.7773 |
| Sensitivity | 151/268 = 0.5634 |
| Specificity | 446/500 = 0.8920 |
| | |
| Positive Predictive Value | 151/205 = 0.7366 |
| | |
| Negative Predictive Value | 446/563 = 0.7922 |

```
          Reference
Prediction   0   1
         0 445 112     This table is slightly different
         1  55 156     than the table used for HW9.

              Accuracy : 0.7826
                95% CI : (0.7517, 0.8112)
   No Information Rate : 0.651
   P-Value [Acc > NIR] : 1.373e-15

                 Kappa : 0.4966

Mcnemar's Test P-Value : 1.468e-05

           Sensitivity : 0.5821
           Specificity : 0.8900
        Pos Pred Value : 0.7393
        Neg Pred Value : 0.7989
            Prevalence : 0.3490
        Detection Rate : 0.2031
  Detection Prevalence : 0.2747
     Balanced Accuracy : 0.7360
```

```
>
> #generate ROC curve   - this information is extra to the Homework
> ROCresult <- roc(pima$test ~ logistic$fitted)
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> plot(ROCresult, legacy.axes = TRUE)
> names(ROCresult)
 [1] "percent"            "sensitivities"      "specificities"
 [4] "thresholds"         "direction"          "cases"
 [7] "controls"           "fun.sesp"           "auc"
[10] "call"               "original.predictor" "original.response"
[13] "predictor"          "response"           "levels"
> ROCresult$auc
Area under the curve: 0.8394
>
> ##-----------------------------------------------------------------#
> dev.off()
null device
          1
>
```

```
#calculate confusion matrix for a given 2 x 2 matrix
classes <- c("0", "1")
truth <- factor(rep(classes, times = c(500,268)),
                levels = (classes))   #enter the reference totals (0,1)
pred <- factor(
            c(
              rep(classes, times = c(446,54)),
              rep(classes, times = c(117,151))),
            levels = (classes))  #enter reference values by column
crosstab <- table(pred, truth)
crosstab
# load Caret package for computing Confusion matrix
library(caret)
confusionMatrix(pred,truth,positive='1')
```