Homework 4

Drew Dahlquist

3/1/2022

1.

2.

(a)

P(Y = 1 | X) = 0.1071429

(b)

P(Y = 1 | X) = 1.0833333

3.

```r
Auto = read.csv("Auto.csv", na.strings="?")
Auto = na.omit(Auto)
Auto.class = Auto[,c(1,2,4,6,7,8)]
Auto.class = Auto.class[Auto.class$origin != '2',]
Auto.class$origin = as.factor(Auto.class$origin)
summary(Auto.class)
```

```
##       mpg          cylinders       horsepower      acceleration
##  Min.   : 9.00   Min.   :3.000   Min.   : 52.00   Min.   : 8.00
##  1st Qu.:16.00   1st Qu.:4.000   1st Qu.: 82.75   1st Qu.:13.50
##  Median :20.90   Median :6.000   Median : 97.00   Median :15.40
##  Mean   :22.57   Mean   :5.747   Mean   :109.49   Mean   :15.28
##  3rd Qu.:28.00   3rd Qu.:8.000   3rd Qu.:140.00   3rd Qu.:17.00
##  Max.   :46.60   Max.   :8.000   Max.   :230.00   Max.   :22.20
##       year       origin
##  Min.   :70.00   1:245
##  1st Qu.:73.00   3: 79
##  Median :76.00
##  Mean   :76.04
##  3rd Qu.:79.00
##  Max.   :82.00
```

(a)

```
glm.fit = glm(origin ~ mpg + cylinders + horsepower + acceleration + year, family=binomial, data=Auto.cl
summary(glm.fit)
```

```
##
## Call:
## glm(formula = origin ~ mpg + cylinders + horsepower + acceleration +
##     year, family = binomial, data = Auto.class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.79504  -0.34097  -0.14154  -0.04997   2.71356
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  10.966194   4.700567   2.333  0.01965 *
## mpg           0.241707   0.051999   4.648 3.35e-06 ***
## cylinders    -1.195342   0.289233  -4.133 3.58e-05 ***
## horsepower    0.028470   0.015685   1.815  0.06952 .
## acceleration -0.003536   0.102960  -0.034  0.97260
## year         -0.198193   0.061297  -3.233  0.00122 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 359.93  on 323  degrees of freedom
## Residual deviance: 207.19  on 318  degrees of freedom
## AIC: 219.19
##
## Number of Fisher Scoring iterations: 7
```

Yes, mpg, cylinders, and year appear to be statistically significant according to the summary.

(b)

The estimated coefficient for cylinders indicates that a 1-unit increase in cylinders will decrease the log-odds of origin by -1.19. Similarly, the coefficient for year indicates that a 1-unit increase in year will decrease the log-odds of origin by -0.19.

(c)

```
glm.probs = predict(glm.fit,type="response")
contrasts(Auto.class$origin)
```

```
##   3
## 1 0
## 3 1
```

```
glm.pred=rep("1", length(Auto.class$origin))
glm.pred[glm.probs > 0.5] = "3"
table(glm.pred, Auto.class$origin)
```

```
##
## glm.pred   1   3
##        1 219  22
##        3  26  57
```

Total fraction of correct predictions is 0.8518519.

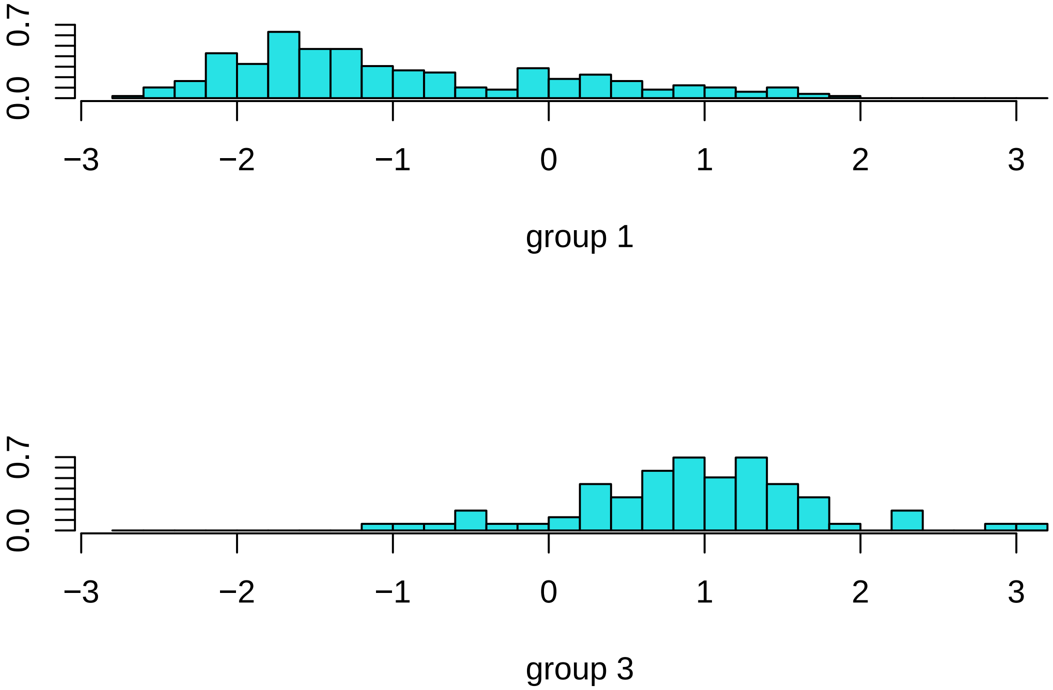
(d)

```
library(MASS)
lda.fit=lda(origin ~ mpg + cylinders + horsepower + acceleration + year, data=Auto.class)
lda.fit
```

```
## Call:
## lda(origin ~ mpg + cylinders + horsepower + acceleration + year,
##     data = Auto.class)
##
## Prior probabilities of groups:
##         1         3
## 0.7561728 0.2438272
##
## Group means:
##        mpg cylinders horsepower acceleration     year
## 1 20.03347  6.277551  119.04898     14.99020 75.59184
## 3 30.45063  4.101266   79.83544     16.17215 77.44304
##
## Coefficients of linear discriminants:
##                       LD1
## mpg           0.166780032
## cylinders    -0.351573318
## horsepower    0.010948130
## acceleration -0.007765445
## year         -0.113510801
```

(e)

```
plot(lda.fit)
```



(f)

```
lda.pred=predict(lda.fit, Auto.class)
lda.class=lda.pred$class
table(lda.class, Auto.class$origin)
```

```
##
## lda.class   1   3
##         1 220  23
##         3  25  56
```

(g)

```
qda.fit=qda(origin ~ mpg + cylinders + horsepower + acceleration + year, data=Auto.class)
qda.fit
```

```
## Call:
## qda(origin ~ mpg + cylinders + horsepower + acceleration + year,
##     data = Auto.class)
##
## Prior probabilities of groups:
##         1         3
## 0.7561728 0.2438272
##
## Group means:
##        mpg cylinders horsepower acceleration     year
## 1 20.03347  6.277551  119.04898     14.99020 75.59184
## 3 30.45063  4.101266   79.83544     16.17215 77.44304
```

(h)

```
qda.pred=predict(qda.fit, Auto.class)
qda.class=qda.pred$class
table(qda.class, Auto.class$origin)
```

```
##
## qda.class   1   3
##         1 194  10
##         3  51  69
```

(i)

Comparison of LDA vs QDA

	LDA	QDA
Accuracy	0.8518519	0.8117284
Sensitivity	0.691358	0.575
Specificity	0.0946502	0.0490196
Precision	0.7088608	0.8734177