Homework 5

Lydia Strebe

July 30, 2019

Problem 1, Part 1

a)

```
Titanic<-read.csv("http://users.stat.umn.edu/~parky/TitanicPartial.csv")
attach(Titanic)
table(Pclass,Survived)

##        Survived
## Pclass   0   1
##      1  64 122
##      2  90  83
##      3 270  85
```

(i) First Class:

```
Prob1=122/(122+64)
Odds1=Prob1/(1-Prob1)
Odds1

## [1] 1.90625
```

Second Class:

```
Prob2=83/(83+90)
Odds2=Prob2/(1-Prob2)
Odds2

## [1] 0.9222222
```

Third Class:

```
Prob3=85/(85+270)
Odds3=Prob3/(1-Prob3)
Odds3

## [1] 0.3148148
```

(ii) Odds ratio:

```
Odds1/Odds3

## [1] 6.055147
```

(iii) Passengers in First Class were over 6 times more likely to survive than passengers in Third Class.

b)

```
m0<-glm(Survived~as.factor(Pclass), family=binomial, data=Titanic)
summary(m0)

##
## Call:
## glm(formula = Survived ~ as.factor(Pclass), family = binomial,
##     data = Titanic)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4607  -0.7399  -0.7399   0.9184   1.6908
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.6451     0.1543   4.180 2.92e-05 ***
## as.factor(Pclass)2  -0.7261     0.2168  -3.350 0.000808 ***
## as.factor(Pclass)3  -1.8009     0.1982  -9.086  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 869.81  on 711  degrees of freedom
## AIC: 875.81
##
## Number of Fisher Scoring iterations: 4
```

(i) Logit Form: log(theta/(1-theta)) = 0.6451-0.7261*x_1-1.8009*x_2

Probability Form: theta=1/(1+e^(-(0.6451-0.7261*x_1-1.8009*x_2)))

(ii) For a person in First Class (the default), the odds of survival were

```
exp(0.6451)

## [1] 1.906178
```

In other words, beta0=log(odds of survival for 1st class)

(iii) The odds ratio for a person in third class is

```
exp(1.8009)

## [1] 6.055095
```

In other words, beta2=-log(odds ratio)

(iv) The estimated probability of surviving for a randomly selected person in First Class is

```
1/(1+exp(-0.6451))

## [1] 0.6559054
```

Problem 1, Part 2

(a) It looks like females in 1st class had the highest odds of survival.

(b)

```
m1=glm(Survived~as.factor(Pclass)+Sex, family=binomial, data=Titanic)
m2=glm(Survived~as.factor(Pclass)+ as.factor(Pclass)*Sex + Age,
family=binomial, data=Titanic)
summary(m1)
```

```
##
## Call:
## glm(formula = Survived ~ as.factor(Pclass) + Sex, family = binomial,
##     data = Titanic)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2029  -0.7371  -0.4556   0.6617   2.1526
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)          2.3338     0.2385   9.784  < 2e-16 ***
## as.factor(Pclass)2  -0.9261     0.2580  -3.590 0.000331 ***
## as.factor(Pclass)3  -1.9748     0.2379  -8.300  < 2e-16 ***
## Sexmale             -2.5721     0.2032 -12.658  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 672.43  on 710  degrees of freedom
## AIC: 680.43
##
## Number of Fisher Scoring iterations: 4
```

```
summary(m2)
```

```
##
## Call:
## glm(formula = Survived ~ as.factor(Pclass) + as.factor(Pclass) *
##     Sex + Age, family = binomial, data = Titanic)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1093  -0.6339  -0.4620   0.3896   2.5299
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 4.909854   0.682782   7.191 6.43e-13 ***
```

```
## as.factor(Pclass)2         -1.160437   0.734046  -1.581  0.11391
## as.factor(Pclass)3         -4.169187   0.648309  -6.431 1.27e-10 ***
## Sexmale                    -3.639225   0.628523  -5.790 7.03e-09 ***
## Age                        -0.041934   0.008184  -5.124 2.99e-07 ***
## as.factor(Pclass)2:Sexmale -0.668839   0.815889  -0.820  0.41235
## as.factor(Pclass)3:Sexmale  2.192651   0.685268   3.200  0.00138 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 613.43  on 707  degrees of freedom
## AIC: 627.43
##
## Number of Fisher Scoring iterations: 6
```

c)

```
pred1=predict(m1, newdata = data.frame(Sex = 'male', Age = 60, Pclass = 3))
1/(1+exp(-pred1))

##          1
## 0.09858161

pred2=predict(m2, newdata = data.frame(Sex = 'male', Age = 60, Pclass = 3))
1/(1+exp(-pred2))

##          1
## 0.03834764
```

The probability of survival is higher using model 1.

d) H_0: beta4, beta5, and beta6 all equal 0

H_1: At least one of these betas does not equal 0

Test statistic:

```
672.43 - 613.43

## [1] 59

pchisq(59, df=3, lower=F)

## [1] 9.613049e-13
```

Since this is much smaller than 0.05, we reject the null hypothesis and conclude that model 1 is not adequate compared to model 2.