MATH523 Assignment 4

Kabilan Sriranjan

March 29, 2018

A12

a)

```
library(aod)
library(VGAM)
```

```
## Loading required package: stats4
```

```
## Loading required package: splines
```

```
library(nnet)
```

A13

a)

```
#A13
I <- c(14, 483, 497, 1008)
notI <- c(1105, 411111, 4624, 157342)
S <- c("Seat belt", "Seat belt", "None", "None")
E <- c("Yes", "No", "Yes", "No")

model.fatal <- glm(cbind(I, notI) ~ S+E, family=binomial)
#likelihood test
pchisq(deviance(model.fatal), df=1, lower.tail=FALSE)
```

```
## [1] 0.09114565
```

```
#pearson chi square test
pchisq(sum(residuals(model.fatal, type="pearson")^2), df=1, lower.tail=FALSE)
```

```
## [1] 0.1098935
```

The p-values we get using both tests are not significant, which means that our model is adequate.

b)

```
summary(model.fatal)
```

```
##
## Call:
## glm(formula = cbind(I, notI) ~ S + E, family = binomial)
##
```

```
## Deviance Residuals:
##       1        2        3        4
## -1.6132   0.3142   0.3256  -0.2165
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.04362    0.03120 -161.65   <2e-16 ***
## SSeat belt  -1.71732    0.05402  -31.79   <2e-16 ***
## EYes         2.79779    0.05526   50.63   <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: 3567.723  on 3  degrees of freedom
## Residual deviance:    2.854  on 1  degrees of freedom
## AIC: 38.039
##
## Number of Fisher Scoring iterations: 3
```

```
beta <- coef(model.fatal)
odds.ratio <- exp(beta)
stdErr <- summary(model.fatal)$coefficients[,2]

lowerCI <- exp(beta-pnorm(0.025)*stdErr)
higherCI <- exp(beta+pnorm(0.025)*stdErr)
#Odds ratio and confidence interval for Seatbelt
odds.ratio[2]
```

```
## SSeat belt
##  0.1795466
```

```
lowerCI[2]
```

```
## SSeat belt
##  0.1746682
```

```
higherCI[2]
```

```
## SSeat belt
##  0.1845612
```

```
#Odds ratio and confidence interval for Ejected
odds.ratio[3]
```

```
##     EYes
## 16.40842
```

```
lowerCI[3]
```

```
##    EYes
## 15.9525
```

```
higherCI[3]
```

```
##     EYes
## 16.87737
```

The odds ratio, r, for Seatbelt can be obtained by

$$
\begin{aligned}
r &= \frac{e^{\beta_0+\beta_1(1)+\beta_2 x_{i2}}}{e^{\beta_0+\beta_1(0)+\beta_2 x_{i2}}} \\
&= e^{\beta_1}
\end{aligned}
$$

And we can do the same for $beta_2$ to get the odds ratio for Ejected. The confidence intervals are obtained through the formula $\widehat{\beta} \pm se(\widehat{\beta}) z_{0.025}$

c)

```
model.fatal2 <- glm(cbind(I, notI) ~ E, family=binomial)
beta2 <- coef(model.fatal2)
odds.ratio2 <- exp(beta2)
stdErr2 <- summary(model.fatal2)$coefficients[,2]
lowerCI2 <- exp(beta2-pnorm(0.025)*stdErr2)
higherCI2 <- exp(beta2+pnorm(0.025)*stdErr2)

#Odds and confidence interval
odds.ratio2[2]
```

```
##     EYes
## 34.00627
```

```
lowerCI2[2]
```

```
##     EYes
## 33.10025
```

```
higherCI2[2]
```

```
##     EYes
## 34.93709
```

We compute the odds ratio and confidence interval for the new model the exact same way as before.

A14

a)

```
#A14
race <- c(1, 1, 0, 0)
gender <- c(1, 0, 1, 0)
heaven.yes <- c(88, 54, 397, 235)
heaven.mb <- c(16, 7, 141, 189)
heaven.no <- c(2, 5, 24, 39)

model.heaven <- multinom(cbind(heaven.yes, heaven.mb, heaven.no) ~ race+gender)
```

```
## # weights:  12 (6 variable)
## initial  value 1315.038910
## iter  10 value 931.782315
```

```
## final  value 930.880924
## converged
```

```
summary(model.heaven)
```

```
## Call:
## multinom(formula = cbind(heaven.yes, heaven.mb, heaven.no) ~
##     race + gender)
##
## Coefficients:
##           (Intercept)       race    gender
## heaven.mb  -0.2633509 -1.1486396 -0.725226
## heaven.no  -1.7942192 -0.6721255 -1.034177
##
## Std. Errors:
##           (Intercept)      race    gender
## heaven.mb  0.09588936 0.2367592 0.1318292
## heaven.no  0.16750483 0.4113510 0.2586682
##
## Residual Deviance: 1861.762
## AIC: 1873.762
```

$log(\pi_{i1}/\pi_{i3})$ is given by

$$\alpha_1 + \beta_1^G x_{i1} + \beta_1^R x_{i2}$$

$log(\pi_{i2}/\pi_{i3})$ is given by

$$\alpha_2 + \beta_2^G x_{i1} + \beta_2^R x_{i2}$$

$log(\pi_{i1}/\pi_{i2})$ is given by

$$log(\pi_{i1}/\pi_{i3}) - log(\pi_{i2}/\pi_{i3})$$

b)

```
model.heavenB <- glm(cbind(heaven.yes, heaven.mb) ~ race+gender, family=binomial)
summary(model.heavenB)
```

```
##
## Call:
## glm(formula = cbind(heaven.yes, heaven.mb) ~ race + gender, family = binomial)
##
## Deviance Residuals:
##       1        2        3        4
## -1.4959   1.7090   0.4840  -0.4869
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.26545    0.09587   2.769  0.00562 **
## race         1.14416    0.23682   4.831 1.36e-06 ***
## gender       0.72245    0.13188   5.478 4.30e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 66.7059  on 3  degrees of freedom
## Residual deviance:  5.6297  on 1  degrees of freedom
## AIC: 32.745
##
## Number of Fisher Scoring iterations: 4
```

```
beta.gender <- coef(model.heavenB)[3]
stderr.gender <- summary(model.heavenB)$coefficients[,2][3]
gender.oddsratio <- exp(beta.gender)
lowerCI.gender <- exp(beta.gender-pnorm(0.025)*stderr.gender)
higherCI.gender <- exp(beta.gender+pnorm(0.025)*stderr.gender)
#Odds ratio and confidence interval
gender.oddsratio
```

```
##   gender
## 2.059466
```

```
lowerCI.gender
```

```
##   gender
## 1.925508
```

```
higherCI.gender
```

```
##   gender
## 2.202742
```

From looking at the odds ratio it appears that males have triple the odds of not believing in heaven than females.

c)

```
model.heavenRace <- glm(cbind(heaven.yes, heaven.no) ~ race, family=binomial)
summary(model.heavenRace)
```

```
##
## Call:
## glm(formula = cbind(heaven.yes, heaven.no) ~ race, family = binomial)
##
## Deviance Residuals:
##      1       2       3       4
##  1.233  -1.238   2.565  -2.773
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.3058     0.1321  17.452   <2e-16 ***
## race          0.7042     0.4091   1.721   0.0852 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 20.816  on 3  degrees of freedom
## Residual deviance: 17.322  on 2  degrees of freedom
```

```
## AIC: 37.622
##
## Number of Fisher Scoring iterations: 4
```

```r
anova(model.heavenB, model.heavenRace, test="Chi")
```

```
## Warning in anova.glmlist(c(list(object), dotargs), dispersion =
## dispersion, : models with response '"cbind(heaven.yes, heaven.no)"' removed
## because response differs from model 1
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: cbind(heaven.yes, heaven.mb)
##
## Terms added sequentially (first to last)
##
##
##        Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                       3     66.706
## race    1   30.746         2     35.960 2.941e-08 ***
## gender  1   30.330         1      5.630 3.645e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can check if gender has an effect on the response when race is controlled by comparing the model with both race and gender to the model with only race. If the test is significant that that means the nested model is not an adequte simplification and so gender does indeed affect the response. Since we do get a significant p-value then that means belief in heaven is not independant of gender when race is controlled for.

A15

If the binary response is the result of an underlying latent variable, this suggests that after some certain threshold, T, the response changes from O to 1. If both groups have similar location then they are both close to T. But because one has significantly higher dispersion than the other, it will have a wide confidence interval that overlaps with the other side of T. This makes it hard to fit a model because if a response is 0 we can't tell if it is truly 0 or actually 1 but the dispersion makes it fall on the wrong side of T.