Stat 536 HW2

Dongyang Wang

2022-10-14

Question 1

```
rm(list = ls())

mental_data <- read.table("mentalimpairment-data.txt",header=TRUE)

alligator_data <- read.table("alligatorfood-data.txt",header=TRUE)
```

Part A

```
attach(mental_data)

#first do not take into consideration the order
library(nnet)
M0 = multinom(Impairment ~ 1)
```

```
## # weights:  8 (3 variable)
## initial  value 55.451774
## final  value 54.521026
## converged
```

```
Ms = multinom(Impairment ~ SES)
```

```
## # weights:  12 (6 variable)
## initial  value 55.451774
## final  value 52.642355
## converged
```

```
Me = multinom(Impairment ~ Events)
```

```
## # weights:  12 (6 variable)
## initial  value 55.451774
## iter  10 value 50.908223
## final  value 50.908198
## converged
```

```
M = multinom(Impairment ~ SES + Events)
```

```
## # weights:  16 (9 variable)
## initial  value 55.451774
## iter  10 value 48.350248
## final  value 48.349131
## converged
```

```r
1-pchisq(Me$deviance-M$deviance,length(coef(M))-length(coef(Me)))
```

```
## [1] 0.1633484
```

```r
1-pchisq(Ms$deviance-M$deviance,length(coef(M))-length(coef(Ms)))
```

```
## [1] 0.03532589
```

```r
detach(mental_data)
```

The above result from MRT shows that the event variable is significant but sec does not seem to matter. This is also supported based on model AIC (Me has lowest AIC).

So the model selected only includes the events variable. The odds are calculated as follows:

```r
exp(1.673461 - 0.2673717 * 7)
```

```
## [1] 0.8202543
```

```r
exp(1.202394 - 0.2835592 * 7)
```

```
## [1] 0.4572521
```

```r
exp(2.453627 - 0.4847294 * 7)
```

```
## [1] 0.3908315
```

That means, compared with being impaired, John is 82.02543% as likely to be mild, 45.72521% as likely to be moderate, 39.08315% as likely to be well.

Part B

```r
attach(mental_data)
#now take the order into account
library(MASS)

Impairment = factor(Impairment,levels=c("Well","Mild","Moderate","Impaired"))

orderM0 = polr(Impairment ~ 1)
orderMs = polr(Impairment ~ SES)
orderMe = polr(Impairment ~ Events)
orderM = polr(Impairment ~ SES + Events)

1-pchisq(orderMe$deviance-orderM$deviance,length(coef(orderM))-length(coef(orderMe)))
```

```
## [1] 0.06405392
```

```r
1-pchisq(orderMs$deviance-orderM$deviance,length(coef(orderM))-length(coef(orderMs)))
```

```
## [1] 0.005293151
```

```r
detach(mental_data)
```

Based on the p-values, SES can be removed from the model – if we allow a significance level of 10% we might keep it as well but for simplicity we keep the 5% significance level. So, we are left with the model with only the events variable.

Assuming logistic error, the results are calculated as follows:

```r
plogis(0.2614334 - 0.28793 * 7)
```

```
## [1] 0.1475338
```

```r
plogis(1.6562752 - 0.28793 * 7) - plogis(0.2614334 - 0.28793 * 7)
```

```
## [1] 0.2636111
```

```r
plogis(2.5876265 - 0.28793 * 7) - plogis(1.6562752 - 0.28793 * 7)
```

```
## [1] 0.2281066
```

```r
1- plogis(2.5876265 - 0.28793 * 7)
```

```
## [1] 0.3607486
```

John is 14.75338% well, 26.36111% mild, 22.81066% moderate, 36.07486% well.

Question 2

```r
attach(alligator_data)

y = cbind(Fish,Invertebrate,Reptile,Bird,Other)


M1 = multinom(y ~ 1)
```

```
## # weights:  10 (4 variable)
## initial  value 352.466903
## final  value 302.181462
## converged
```

```r
M2 = multinom(y ~ Lake)
```

```
## # weights:  25 (16 variable)
## initial  value 352.466903
## iter  10 value 281.030560
## iter  20 value 280.583926
## final  value 280.583844
## converged
```

```r
M3 = multinom(y ~ Gender)
```

```
## # weights:  15 (8 variable)
## initial  value 352.466903
## iter  10 value 301.192714
## final  value 301.129428
## converged
```

```r
M4 = multinom(y ~ Size)
```

```
## # weights:  15 (8 variable)
## initial  value 352.466903
## iter  10 value 294.689203
## final  value 294.606678
## converged
```

```r
M5 = multinom(y ~ Lake + Gender)
```

```
## # weights:  30 (20 variable)
## initial  value 352.466903
```

```
## iter  10 value 279.421076
## iter  20 value 277.733691
## final  value 277.732884
## converged
```

```r
M6 = multinom(y ~ Lake + Size)
```

```
## # weights:  30 (20 variable)
## initial  value 352.466903
## iter  10 value 270.844027
## iter  20 value 270.041364
## final  value 270.040140
## converged
```

```r
M7 = multinom(y ~ Gender + Size)
```

```
## # weights:  20 (12 variable)
## initial  value 352.466903
## iter  10 value 294.854746
## final  value 294.091727
## converged
```

```r
M8 = multinom(y ~ Lake+Gender+Size)
```

```
## # weights:  35 (24 variable)
## initial  value 352.466903
## iter  10 value 271.274792
## iter  20 value 268.935538
## final  value 268.932740
## converged
```

```r
AIC(M1)
```

```
## [1] 612.3629
```

```r
AIC(M2)
```

```
## [1] 593.1677
```

```r
AIC(M3)
```

```
## [1] 618.2589
```

```r
AIC(M4)
```

```
## [1] 605.2134
```

```r
AIC(M5)
```

```
## [1] 595.4658
```

```r
AIC(M6)
```

```
## [1] 580.0803
```

```r
AIC(M7)
```

```
## [1] 612.1835
```

```r
AIC(M8)
```

```
## [1] 585.8655
```

```
1-pchisq(M6$deviance-M8$deviance,length(coef(M8))-length(coef(M6)))
```

```
## [1] 0.6963208
```

```
detach(alligator_data)
```

Based on AIC and LRT, Lake and Size seem to be the only variables that are significant. Compared with alligators from Lake George with small size, the model would predict as follows:

```
exp(-3.66588368 + 2.935296 + 0.3513836)
```

```
## [1] 0.6844059
```

it is 68.44059% as likely to be reptile eater as to be a fish eater.