Hw8

Mohammad Mottahedi

March 20, 2016

```
set.seed(1234)
setwd('~/Dropbox/spring2016/discretDataAnalysis/hw_solutions/hw8/')
library(VGAM)
```

```
## Loading required package: stats4

## Loading required package: splines

##
## Attaching package: 'VGAM'

## The following object is masked from '.env':
##
##     s
```

```
library(MASS)
```

Problem 1

part a

```
gator = read.table("gator.txt",header=T)
gator$Size = factor(gator$Size,levels=levels(gator$Size)[2:1])
totaln=sum(gator[1:16,5:9]) ## total sample size
rown=c(1:16)
for (i in 1:16) {
rown[i]=sum(gator[i,5:9])
rown
}

##sets Hancock as the baseline level
contrasts(gator$Lake)=contr.treatment(levels(gator$Lake),base=2)
#contrasts(gator$Lake)

##sets "small" as the refernce level
contrasts(gator$Size)=contr.treatment(levels(gator$Size),base=2)
#contrasts(gator$Size)


##sets male as the reference level
contrasts(gator$Gender)=contr.treatment(levels(gator$Gender),base=2)
#contrasts(gator$Gender)

fit = vglm(cbind(Bird,Invertebrate,Reptile,Other,Fish)~Lake+Size, data=gator, family=multinomial)
summary(fit)
```

```
##
## Call:
## vglm(formula = cbind(Bird, Invertebrate, Reptile, Other, Fish) ~
##     Lake + Size, family = multinomial, data = gator)
##
## Pearson residuals:
##                         Min      1Q  Median     3Q   Max
## log(mu[,1]/mu[,5]) -0.9873 -0.5082 -0.1144 0.2373 3.994
## log(mu[,2]/mu[,5]) -1.3716 -0.4379 -0.0248 0.2436 1.995
## log(mu[,3]/mu[,5]) -0.8298 -0.5850 -0.2309 0.2225 2.237
## log(mu[,4]/mu[,5]) -1.5873 -0.3189 -0.0159 1.0330 1.413
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept):1   -2.0286     0.5581  -3.635 0.000278 ***
## (Intercept):2   -1.7492     0.5392  -3.244 0.001178 **
## (Intercept):3   -2.4230     0.6436  -3.765 0.000167 ***
## (Intercept):4   -0.7465     0.3520  -2.121 0.033928 *
## Lakegeorge:1    -0.6951     0.7813  -0.890 0.373608
## Lakegeorge:2     1.6584     0.6129   2.706 0.006813 **
## Lakegeorge:3    -1.2428     1.1854  -1.048 0.294461
## Lakegeorge:4    -0.8262     0.5575  -1.482 0.138378
## Lakeoklawaha:1  -1.3483     1.1633  -1.159 0.246453
## Lakeoklawaha:2   2.5956     0.6597   3.934 8.34e-05 ***
## Lakeoklawaha:3   1.2161     0.7860   1.547 0.121823
## Lakeoklawaha:4  -0.8205     0.7296  -1.125 0.260753
## Laketrafford:1   0.3926     0.7818   0.502 0.615487
## Laketrafford:2   2.7803     0.6712   4.142 3.44e-05 ***
## Laketrafford:3   1.6925     0.7804   2.169 0.030113 *
## Laketrafford:4   0.6902     0.5597   1.233 0.217512
## Size>2.3:1       0.6307     0.6425   0.982 0.326291
## Size>2.3:2      -1.4582     0.3959  -3.683 0.000231 ***
## Size>2.3:3       0.3513     0.5800   0.606 0.544785
## Size>2.3:4      -0.3316     0.4483  -0.740 0.459511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors:  4
##
## Names of linear predictors:
## log(mu[,1]/mu[,5]), log(mu[,2]/mu[,5]), log(mu[,3]/mu[,5]), log(mu[,4]/mu[,5])
##
## Dispersion Parameter for multinomial family:   1
##
## Residual deviance: 52.4785 on 44 degrees of freedom
##
## Log-likelihood: -74.4295 on 44 degrees of freedom
##
## Number of iterations: 5
##
## Reference group is level  5  of the response
```

The overal fit of the model can be evaluiated using residual deviance 52.4 with 44 degrees of freedom (p-value: 0.1803193). Overal the fit of the model is not good.

model equations:

$log\frac{\pi_{food,lsg}}{\pi_{F,lsg}} = \alpha_{food} + \beta^{lake}_{food,l} + \beta^{size}_{food,s}$

$log\frac{\pi_{bird,lsg}}{\pi_{F,lsg}} = -2.03 - 1.35 Oklawaha + 0.39 Trafford - 0.695 George + 0.6307 Large$

$log\frac{\pi_{Invertebrate,lsg}}{\pi_{F,lsg}} = -1.7492 + 2.5956 Oklawaha + 2.7803 Trafford - 0.695 George - 1.4582 Large$

$log\frac{\pi_{Invertebrate,lsg}}{\pi_{F,lsg}} = -2.4230 + 1.2161 Oklawaha + 1.6925 Trafford - 1.2428 George + 0.3513 Large$

$log\frac{\pi_{Invertebrate,lsg}}{\pi_{F,lsg}} = -0.7465 - 0.8205 Oklawaha + 0.6902 Trafford - 0.8262 George - 0.3316 Large$

part b

```
denom = 1 + exp(-2.03 - 1.35  + 0.6307 ) +
    exp(-1.7492 + 2.5956 - 1.4582 ) +
    exp(-2.4230 + 1.2161 + 0.3513 ) +
    exp(-0.7465 - 0.8205 - 0.3316 )
```

$P(fish|Large, ...) = 0.4584732$

```
denom = 1 + exp(-2.03 - 1.35) +
    exp(-1.7492 + 2.5956) +
    exp(-2.4230 + 1.2161) +
    exp(-0.7465 - 0.8205)
```

$P(fish|small, ...) = 0.2581924$

probability of fish being the main food source is 1.7 times (46%) for large aligators compared to small aligators (0.26)

part c

```
gator = read.table("alligator.dat",col.names = c('lake', 'sex', 'size', 'food','count'))
#gator$size = factor(gator$size,levels=c("small", "large"))
gator$food = factor(gator$food, levels = c('fish', 'invertebrate', 'reptile', 'bird', 'other'))
result = polr( food ~ size + lake, weights=count, data=gator)
null_model = polr( food ~ 1, weights=count, data=gator)
sat_model = polr(food ~ size + lake + sex + size:lake + size:sex + sex:lake, weights=count, data=gator )

summary(result)
```

```
##
## Re-fitting to get Hessian

## Call:
## polr(formula = food ~ size + lake, data = gator, weights = count)
##
## Coefficients:
##                 Value Std. Error t value
## sizesmall     0.02694     0.3394 0.07939
## lakeHancock   0.87186     0.4587 1.90062
```

```
## lakeOklawaha 0.46396     0.5200 0.89219
## lakeTrafford 1.43344     0.4750 3.01774
##
## Intercepts:
##                      Value  Std. Error t value
## fish|invertebrate    1.1239 0.4029     2.7894
## invertebrate|reptile 1.1239 0.4029     2.7894
## reptile|bird         1.6939 0.4144     4.0878
## bird|other           2.1600 0.4273     5.0552
##
## Residual Deviance: 334.5842
## AIC: 350.5842
## (16 observations deleted due to missingness)
```

```r
# comparing with null model
anova(result, null_model)
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: food
##         Model Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
## 1           1       154   345.2587
## 2 size + lake       150   334.5842 1 vs 2     4 10.67453 0.03047597
```

```r
# comparing with sat model
anova(result, sat_model)
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: food
##                                                 Model Resid. df Resid. Dev
## 1                                         size + lake       150   334.5842
## 2 size + lake + sex + size:lake + size:sex + sex:lake       142   327.1430
##     Test    Df LR stat.   Pr(Chi)
## 1
## 2 1 vs 2     8 7.441164 0.4898658
```

```r
 pchisq(deviance(result), df.residual(result), lower.tail = F)
```

```
## [1] 4.108487e-16
```

The model is significant compared to both null and saturated model which shows that this model at leat has one non-zero coefficeint. The p-value for the current model is close to zero so the model is a good fit.

The effect of size is oposite of what we found in part b. the probability of eating fish is greater for smaller aligators.

Problem 2

part a

We have to fit $r-1$ equation which is equal to 3. The 3 parameter coefficients are the same and the 3 interceprt are different.

part b

x_1 : decrease

x_2 : increase

x_3 : increase

part c

x_1 : much less

x_2 : not at all true

x_3 : strongly disagree

Problem 3

$logit[P(Y \leq j|x_1)] - logit[P(Y \leq j|x_2)]$

$= log\frac{P(Y \leq j|x_1)/P(Y > j|x_1)}{P(Y \leq j|x_2)/P(Y > j|x_2)}$

$log[(1-\alpha-\beta x_2)(\alpha+\beta x_1)-(1-\alpha-\beta x_1)(\alpha+\beta x_2)]$

$= \beta(x_1 - x_2)$

if $x_1 - x_2 = 1$

$= \beta$

part b

from part a:

$log\frac{P(Y \leq j|x_1)/P(Y > j|x_1)}{P(Y \leq j|x_2)/P(Y > j|x_2)} = \beta(x_1 - x2)$

the odds of being below $\leq j$ at x_1 is $e^{\beta(x_1 - x_2)}$ times the odds at x_2.

for x=1,2 and j=1,2

	j1	j2
x1	p(y <= 1 / x1)	p(y >= 2 / x1)
x2	p(y <= 1 / x2)	p(y >= 2 / x2)