HW6_1830

Yue Zhang

2024-12-01

#1

```r
afterlife = read.table("http://www.stat.ufl.edu/~aa/cat/data/Afterlife.dat",
    header = TRUE)
afterlife_model = vglm(cbind(yes, undecided, no) ~ gender + race,
    family = multinomial, data = afterlife)
summary(afterlife_model)
```

```
## Call:
## vglm(formula = cbind(yes, undecided, no) ~ gender + race, family = multinomial,
##     data = afterlife)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1   1.3016     0.2265   5.747  9.1e-09 ***
## (Intercept):2  -0.6529     0.3405  -1.918   0.0551 .
## gendermale:1   -0.4186     0.1713  -2.444   0.0145 *
## gendermale:2   -0.1051     0.2465  -0.426   0.6700
## racewhite:1     0.3418     0.2370   1.442   0.1493
## racewhite:2     0.2710     0.3541   0.765   0.4442
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 0.8539 on 2 degrees of freedom
##
## Log-likelihood: -19.7324 on 2 degrees of freedom
##
## Number of Fisher scoring iterations: 3
##
## No Hauck-Donner effect found in any of the estimates
##
##
## Reference group is level  3  of the response
```

```r
# (a)
exp(-0.105)
```

```
## [1] 0.9003245
```

```r
# (b)
exp(-0.429 - (-0.105))
```

```
## [1] 0.7232502
```

(a) The odds ratio for undecided and no pair is 0.9 which means that men are 0.9 times likely to response undecided rather than no than females.
(b) The odds ratio for yes and undecided pair is 1.311 which means that men are 1.311 times likely to response yes rather than undecided than females.

#2

```r
happy = data.frame(income = 1:3, y1 = c(6, 6, 6), y2 = c(43,
    113, 57), y3 = c(75, 178, 117))

happy_model = vglm(cbind(y1, y2, y3) ~ income, data = happy,
    family = cumulative(parallel = TRUE))
summary(happy_model)
```

```
## Call:
## vglm(formula = cbind(y1, y2, y3) ~ income, family = cumulative(parallel = TRUE),
##     data = happy)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  -3.2466     0.3404  -9.537   <2e-16 ***
## (Intercept):2  -0.2378     0.2592  -0.917    0.359
## income         -0.1117     0.1179  -0.948    0.343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])
##
## Residual deviance: 3.2472 on 3 degrees of freedom
##
## Log-likelihood: -15.4146 on 3 degrees of freedom
##
## Number of Fisher scoring iterations: 4
##
## No Hauck-Donner effect found in any of the estimates
##
##
## Exponentiated coefficients:
##    income
## 0.8942746
```

```r
happy_model2 = vglm(cbind(y1, y2, y3) ~ income, data = happy,
    family = cumulative(parallel = FALSE))
summary(happy_model2)
```

```
## Call:
```

```
## vglm(formula = cbind(y1, y2, y3) ~ income, family = cumulative(parallel = FALSE),
##     data = happy)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  -3.0693     0.7170  -4.281 1.86e-05 ***
## (Intercept):2  -0.2467     0.2614  -0.944    0.345
## income:1       -0.1995     0.3380  -0.590    0.555
## income:2       -0.1075     0.1190  -0.904    0.366
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])
##
## Residual deviance: 3.175 on 2 degrees of freedom
##
## Log-likelihood: -15.3785 on 2 degrees of freedom
##
## Number of Fisher scoring iterations: 4
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1'
##
##
## Exponentiated coefficients:
##  income:1  income:2
## 0.8191538 0.8980712
```

```r
lrtest(happy_model, happy_model2)
```

```
## Likelihood ratio test
##
## Model 1: cbind(y1, y2, y3) ~ income
## Model 2: cbind(y1, y2, y3) ~ income
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   3 -15.415
## 2   2 -15.379 -1 0.0722     0.7882
```

The p-value for the likelihood ratio test is 0.788 which is larger than 0.05. Therefore, we cannot reject the null hypothesis. This indicates that the proportional odds assumption holds and the simpler proportional odds model is appropriate for analyzing the effect of income on happiness. Thus, we can fit an ordinal regression model here.

#3

```r
alligator = read.csv("E:/Biostat/Biostatistics/PH 1830/alligator.csv")
alligator_model = vglm(food ~ length, family = multinomial, data = alligator)
summary(alligator_model)
```

```
## Call:
## vglm(formula = food ~ length, family = multinomial, data = alligator)
##
## Coefficients:
```

```
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept):1   1.6177     1.3073   1.237  0.21591
## (Intercept):2   5.6974     1.7937   3.176  0.00149 **
## length:1       -0.1101     0.5171  -0.213  0.83137
## length:2       -2.4654     0.8996      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 98.3412 on 114 degrees of freedom
##
## Log-likelihood: -49.1706 on 114 degrees of freedom
##
## Number of Fisher scoring iterations: 5
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## 'length:2'
##
##
## Reference group is level  3  of the response
```

```r
1 - exp(-0.11)
```

```
## [1] 0.1041659
```

```r
1 - exp(-2.465)
```

```
## [1] 0.9149912
```

```r
exp(2.355)
```

```
## [1] 10.53813
```

The ML prediction equations are: $logit(\frac{\hat{\pi}_1}{\hat{\pi}_3}) = 1.618 - 0.110 * x$, $logit(\frac{\hat{\pi}_2}{\hat{\pi}_3}) = 5.698 - 2.465 * x$, $logit(\frac{\hat{\pi}_1}{\hat{\pi}_2}) = -4.080 + 2.355 * x$ For one unit increases in length, the odds of choosing fish rather than other decrease by 10.4%. For one unit increases in length, the odds of choosing invertebrate rather than other decrease by 91.5%. For one unit increases in length, the odds of choosing fish is 10.538 times of choosing invertebrate.