ordinal logistic regression model demo

高嘉妤

2024-10-10

Table of contents

read data

```
library(table1)
```

```
Attaching package: 'table1'
```

```
The following objects are masked from 'package:base':

    units, units<-
```

```
#grouped data file
Polviews <- read.table("http://www.stat.ufl.edu/~aa/cat/data/Polviews.dat",header=TRUE)
#ungrouped data file
Polviews2 <- read.table("http://www.stat.ufl.edu/~aa/cat/data/Polviews2.dat",header=TRUE)
table1(~gender+party|ideology,data=Polviews2)
```

```
Warning in table1.formula(~gender + party | ideology, data = Polviews2): Terms
to the right of '|' in formula 'x' define table columns and are expected to be
factors with meaningful labels.
```

	1	2	3	4	5	Overall
	(N=45)	(N=184)	(N=158)	(N=203)	(N=71)	(N=661)
gender						
female	25 (55.6%)	110 (59.8%)	101 (63.9%)	111 (54.7%)	36 (50.7%)	383 (57.9%)
male	20 (44.4%)	74 (40.2%)	57 (36.1%)	92 (45.3%)	35 (49.3%)	278 (42.1%)
party						
dem	45 (100%)	178 (96.7%)	129 (81.6%)	48 (23.6%)	7 (9.9%)	407 (61.6%)
repub	0 (0%)	6 (3.3%)	29 (18.4%)	155 (76.4%)	64 (90.1%)	254 (38.4%)

cumulative logit model

```
library(VGAM)
```

```
Loading required package: stats4
```

```
Loading required package: splines
```

```
fit <- vglm(cbind(y1,y2,y3,y4,y5)~party+gender,family=cumulative(parallel=TRUE),data=Polviews)
summary(fit)
```

```
Call:
vglm(formula = cbind(y1, y2, y3, y4, y5) ~ party + gender, family = cumulative(parallel = TRUE),
    data = Polviews)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -2.12233    0.16875 -12.577   <2e-16 ***
(Intercept):2  0.16892    0.11481   1.471    0.141
(Intercept):3  1.85716    0.15103  12.297   <2e-16 ***
(Intercept):4  4.65005    0.23496  19.791   <2e-16 ***
partyrepub    -3.63366    0.21785 -16.680   <2e-16 ***
gendermale     0.04731    0.14955   0.316    0.752
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),
logitlink(P[Y<=3]), logitlink(P[Y<=4])

Residual deviance: 9.8072 on 10 degrees of freedom

Log-likelihood: -35.2032 on 10 degrees of freedom

Number of Fisher scoring iterations: 4

No Hauck-Donner effect found in any of the estimates


Exponentiated coefficients:
partyrepub gendermale
0.02641936 1.04844945
```

For any fixed j, the estimated odds that a Republician's response is in the liberal direction rather than the conservation direction equal exp(β1) = exp(-3.634) = 0.0026 times the estimated odds for Democrats. In other words, The estimated odds that a Republician's response is in the conservative direction rather than the liberal direction equal exp(-β1) = exp(3.634) = 37.9 times the estimated odds for Democrats. We can conclude that strong Republicans tend to be much more conservative than strong Democrats.

```
attach(Polviews)
gendername <- c("female","male","female","male")
partyname <- c("dem","rep","dem","rep")
data.frame(gendername,partyname,fitted(fit))
```

```
  gendername partyname          y1         y2        y3        y4          y5
1     female       dem 0.106945695 0.43518307 0.3228363 0.1255644 0.009470552
2       male       rep 0.003153813 0.02717842 0.1144033 0.5895341 0.265730342
```

```
3     female       dem 0.111549168 0.44229817 0.3165490 0.1205668 0.009036867
4       male       rep 0.003306108 0.02844904 0.1189361 0.5927070 0.256601745
```

The data frame can help us to understand the effects of the explanatory variables. For each political party affiliation, the estimated distributions are very similar for females and males. The most common response is 'slightly liberal' for Democrats, who are very likely to be in category 3 or below, and 'slightly conseervative' for Republicans, who are very likely to be in category 4 or 5.

cumulative logit model without proportional odds

```
fit2 <- vglm(cbind(y1,y2,y3,y4,y5)~party+gender,family=cumulative,data=Polviews)
summary(fit2)
```

```
Call:
vglm(formula = cbind(y1, y2, y3, y4, y5) ~ party + gender, family = cumulative,
    data = Polviews)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept):1   -2.17521    0.20941 -10.387  < 2e-16 ***
(Intercept):2    0.12173    0.12476   0.976    0.329
(Intercept):3    1.88810    0.17043  11.078  < 2e-16 ***
(Intercept):4    4.10365    0.39770  10.318  < 2e-16 ***
partyrepub:1   -20.76294 3458.04727      NA       NA
partyrepub:2    -3.94288    0.42696  -9.235  < 2e-16 ***
partyrepub:3    -3.68095    0.23285 -15.808  < 2e-16 ***
partyrepub:4    -2.94499    0.40785  -7.221 5.17e-13 ***
gendermale:1     0.21835    0.31762   0.687    0.492
gendermale:2     0.18343    0.19352   0.948    0.343
gendermale:3    -0.08638    0.22198  -0.389    0.697
gendermale:4    -0.14633    0.26939  -0.543    0.587
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),
logitlink(P[Y<=3]), logitlink(P[Y<=4])

Residual deviance: 3.5861 on 4 degrees of freedom

Log-likelihood: -32.0927 on 4 degrees of freedom

Number of Fisher scoring iterations: 17

Warning: Hauck-Donner effect detected in the following estimate(s):
'(Intercept):4', 'partyrepub:1', 'partyrepub:2'


Exponentiated coefficients:
partyrepub:1 partyrepub:2 partyrepub:3 partyrepub:4 gendermale:1 gendermale:2
9.611007e-10 1.939224e-02 2.519906e-02 5.260246e-02 1.244018e+00 1.201330e+00
gendermale:3 gendermale:4
9.172418e-01 8.638714e-01
```

The data frame can help us to understand the effects of the explanatory variables. For each political party affiliation,

the estimated distributions are very similar for females and males. The most common response is 'slightly liberal' for Democrats, who are very likely to be in category 3 or below, and 'slightly conseervative' for Republicans, who are very likely to be in category 4 or 5.

cumulative logit model without proportional odds

```
fit2 <- vglm(cbind(y1,y2,y3,y4,y5)~party+gender,family=cumulative,data=Polviews)
summary(fit2)
```

```
Call:
vglm(formula = cbind(y1, y2, y3, y4, y5) ~ party + gender, family = cumulative,
    data = Polviews)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept):1   -2.17521    0.20941 -10.387  < 2e-16 ***
(Intercept):2    0.12173    0.12476   0.976    0.329
(Intercept):3    1.88810    0.17043  11.078  < 2e-16 ***
(Intercept):4    4.10365    0.39770  10.318  < 2e-16 ***
partyrepub:1   -20.76294 3458.04727      NA       NA
partyrepub:2    -3.94288    0.42696  -9.235  < 2e-16 ***
partyrepub:3    -3.68095    0.23285 -15.808  < 2e-16 ***
partyrepub:4    -2.94499    0.40785  -7.221 5.17e-13 ***
gendermale:1     0.21835    0.31762   0.687    0.492
gendermale:2     0.18343    0.19352   0.948    0.343
gendermale:3    -0.08638    0.22198  -0.389    0.697
gendermale:4    -0.14633    0.26939  -0.543    0.587
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),
logitlink(P[Y<=3]), logitlink(P[Y<=4])

Residual deviance: 3.5861 on 4 degrees of freedom

Log-likelihood: -32.0927 on 4 degrees of freedom

Number of Fisher scoring iterations: 17

Warning: Hauck-Donner effect detected in the following estimate(s):
'(Intercept):4', 'partyrepub:1', 'partyrepub:2'


Exponentiated coefficients:
partyrepub:1 partyrepub:2 partyrepub:3 partyrepub:4 gendermale:1 gendermale:2
9.611007e-10 1.939224e-02 2.519906e-02 5.260246e-02 1.244018e+00 1.201330e+00
gendermale:3 gendermale:4
9.172418e-01 8.638714e-01
```

The political party estimate of β1hat = 2.03 means that for the normal latent variable model, with higher y* values representing greater convervatism, the estimated mean political ideology for Republicans is 2.03 higher than the estimated mean for Democrats. This difference is relative to a residual standard deviation of 1.0 for the normal latent response. With an arbitary standard deviation, we estimate that the two groups have means that differ by 2.03 standard deviations,

This is an extremely large effect.

Adjacent-Categories logits

```
fit_acat <- vglm(cbind(y1,y2,y3,y4,y5)~party+gender,family=acat(parallel=TRUE,reverse=TRUE),data=Polview
summary(fit_acat)
```

```
Call:
vglm(formula = cbind(y1, y2, y3, y4, y5) ~ party + gender, family = acat(parallel = TRUE,
    reverse = TRUE), data = Polviews)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -1.38707    0.17079  -8.121 4.61e-16 ***
(Intercept):2  0.36218    0.11749   3.083  0.00205 **
(Intercept):3  0.77529    0.14720   5.267 1.39e-07 ***
(Intercept):4  2.99240    0.23132  12.936  < 2e-16 ***
partyrepub    -2.23478    0.16841 -13.270  < 2e-16 ***
gendermale     0.01212    0.09661   0.125  0.90016
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: loglink(P[Y=1]/P[Y=2]), loglink(P[Y=2]/P[Y=3]),
loglink(P[Y=3]/P[Y=4]), loglink(P[Y=4]/P[Y=5])

Residual deviance: 13.4665 on 10 degrees of freedom

Log-likelihood: -37.0329 on 10 degrees of freedom

Number of Fisher scoring iterations: 4

Warning: Hauck-Donner effect detected in the following estimate(s):
'partyrepub'


Exponentiated coefficients:
partyrepub gendermale
 0.1070151  1.0121943
```

The estimated odds that a Democrats's political ideology is in category j instead of j+1 are exp(-β1)=9.34 times the estimated odds for Republicans.

sequential logits

```
fit_sratio <- vglm(cbind(y1,y2,y3,y4,y5)~party+gender,family=sratio(parallel=TRUE),data=Polviews)
summary(fit_sratio)
```

```
Call:
vglm(formula = cbind(y1, y2, y3, y4, y5) ~ party + gender, family = sratio(parallel = TRUE),
    data = Polviews)

Coefficients:
```

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -2.12723    0.16537 -12.863  < 2e-16 ***
(Intercept):2 -0.09593    0.11297  -0.849    0.396
(Intercept):3  0.88414    0.15341   5.763 8.25e-09 ***
(Intercept):4  3.71958    0.24101  15.433  < 2e-16 ***
partyrepub    -2.96103    0.19164 -15.451  < 2e-16 ***
gendermale     0.00667    0.12793   0.052    0.958
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: logitlink(P[Y=1|Y>=1]), logitlink(P[Y=2|Y>=2]),
logitlink(P[Y=3|Y>=3]), logitlink(P[Y=4|Y>=4])

Residual deviance: 25.9331 on 10 degrees of freedom

Log-likelihood: -43.2662 on 10 degrees of freedom

Number of Fisher scoring iterations: 5

No Hauck-Donner effect found in any of the estimates


Exponentiated coefficients:
partyrepub gendermale
0.05176572 1.00669183
```

For any fixed j, the estimated odds that a Republican's response is in category j rather than higher categories is exp(β1)=0.052 times the estimated odds for Democrats.