

MA678 Homework 6

JingjianGao

11/8/2022

```
knitr::opts_chunk$set(echo = TRUE,out.width="0.9\\linewidth",dev="pdf",fig.align = 'center')
library(ggplot2)
library(knitr)
library(arm)
```

```
## Loading required package: MASS
## Loading required package: Matrix
## Loading required package: lme4
##
## arm (Version 1.13-1, built: 2022-8-25)
## Working directory is /Users/billg/Desktop/MA-678-Homework/MA678-HW6
```

```
library(data.table)
library(foreign)
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:arm':
##
##      logit
```

```
library(faraway)
```

```
##
## Attaching package: 'faraway'
## The following objects are masked from 'package:car':
##
##      logit, vif
```

```
## The following objects are masked from 'package:arm':
##
##      fround, logit, pfround
```

```
library(nnet)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
## The following objects are masked from 'package:data.table':
```

```
##
##      dcast, melt
library(VGAM)

## Loading required package: stats4
## Loading required package: splines
##
## Attaching package: 'VGAM'
## The following objects are masked from 'package:faraway':
##
##      hormone, logit, pneumo, prplot
## The following object is masked from 'package:car':
##
##      logit
## The following object is masked from 'package:arm':
##
##      logit
```

Multinomial logit

Using the individual-level survey data from the 2000 National Election Study (data in folder NES), predict party identification (which is on a five-point scale) using ideology and demographics with an ordered multinomial logit model.

1. Summarize the parameter estimates numerically and also graphically.

```
nes <- read.dta("/Users/billg/Desktop/MA-678-Homework/MA678-HW6/nes5200_processed_voters_realideo.dta")
partyid3_b <- factor(nes$partyid3_b)
ideo <- factor(nes$ideo)
income <- factor(nes$income)
educ3 <- factor(nes$educ3)
Reg <- vglm(partyid3_b ~ ideo+income+educ3,family=multinomial,data=nes)
summary(Reg)
```

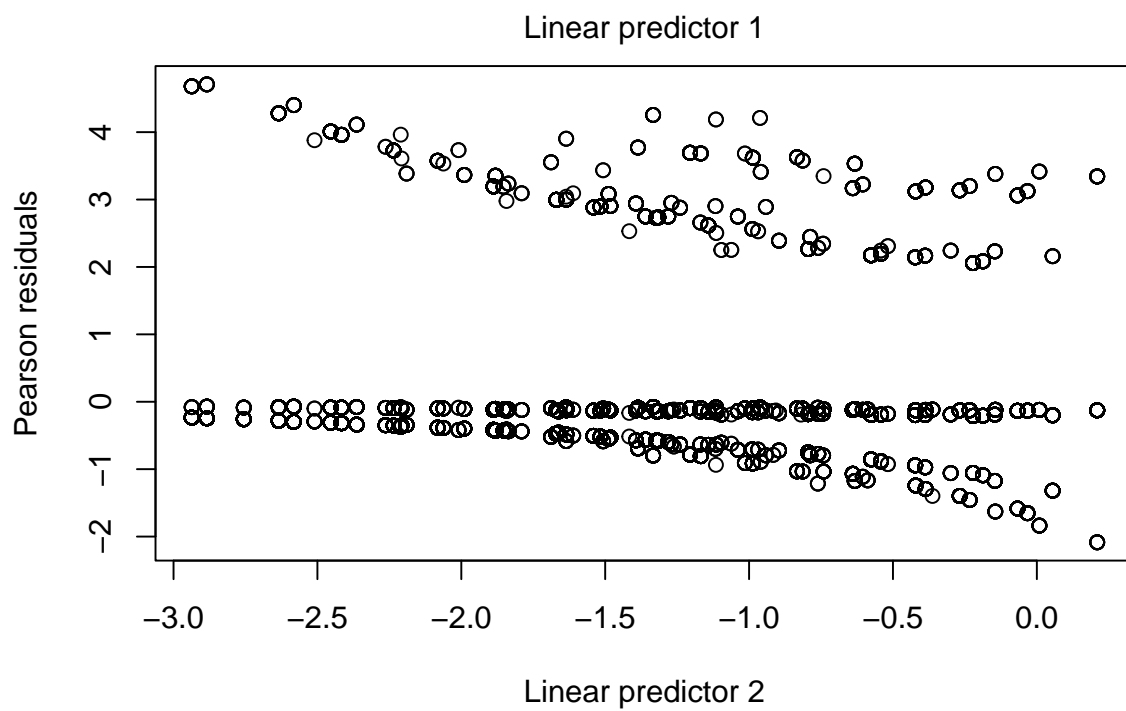
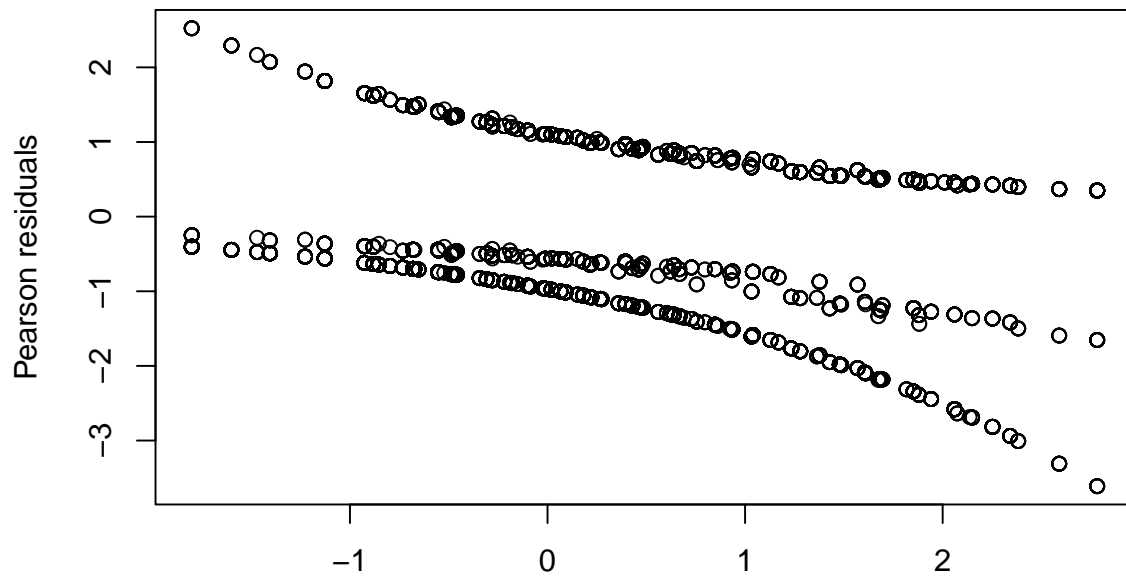
```
##
## Call:
## vglm(formula = partyid3_b ~ ideo + income + educ3, family = multinomial,
##      data = nes)
##
## Coefficients:
##                                     Estimate Std. Error z value
## (Intercept):1                      2.78210     0.11829  23.520
## (Intercept):2                      0.20966     0.17596   1.192
## ideo3. moderate ('middle of the road'):1 -1.21282     0.06446 -18.814
## ideo3. moderate ('middle of the road'):2 -0.15483     0.09960  -1.554
## ideo5. conservative:1                -2.16052     0.05490 -39.353
## ideo5. conservative:2                -1.24858     0.09178 -13.604
## income2. 17 to 33 percentile:1        -0.19233     0.08041  -2.392
## income2. 17 to 33 percentile:2        -0.20086     0.12224  -1.643
## income3. 34 to 67 percentile:1        -0.39980     0.07119  -5.616
## income3. 34 to 67 percentile:2        -0.35413     0.10876  -3.256
## income4. 68 to 95 percentile:1        -0.64587     0.07419  -8.706
## income4. 68 to 95 percentile:2        -0.57239     0.11578  -4.944
```

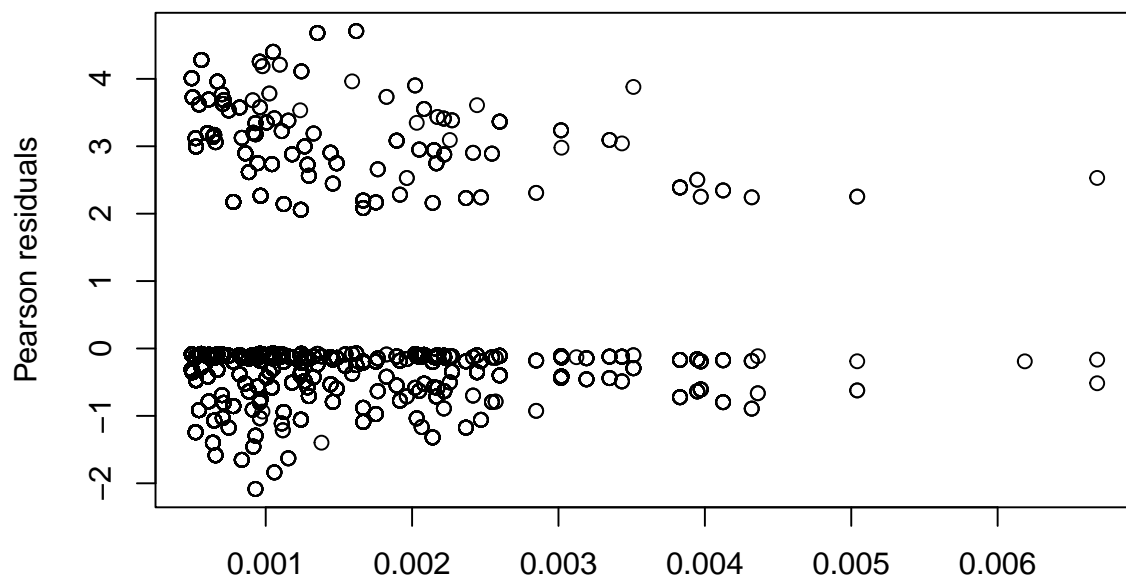
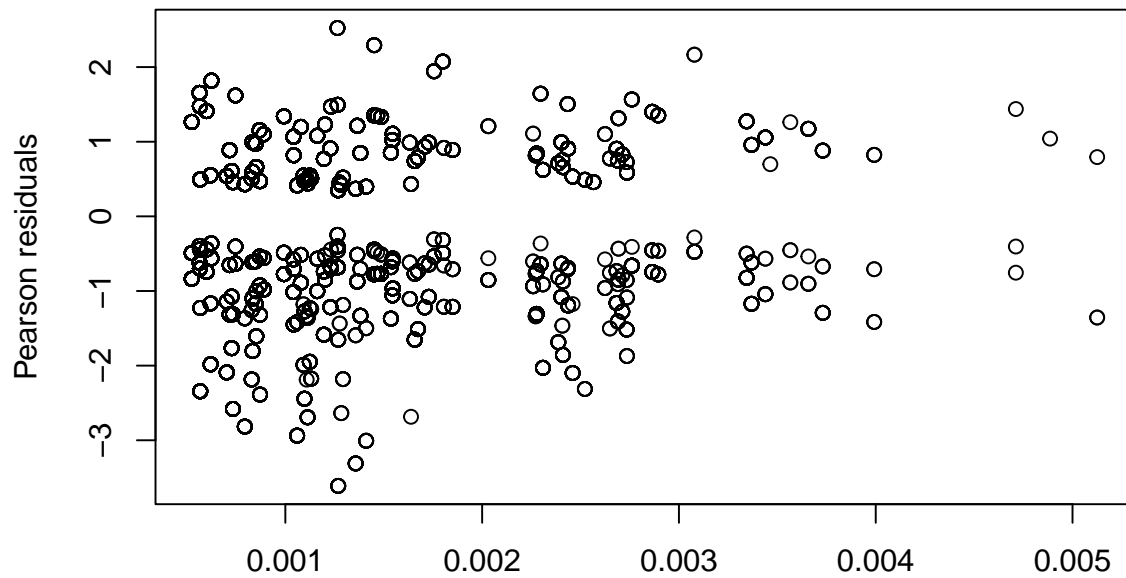
```

## income5. 96 to 100 percentile:1      -1.31898      0.11640 -11.332
## income5. 96 to 100 percentile:2      -0.87449      0.19204  -4.554
## educ32. 9-12 grades ('high school'), no diplo:1 -0.44079      0.12282  -3.589
## educ32. 9-12 grades ('high school'), no diplo:2 -0.24217      0.18191  -1.331
## educ33. 12 grades, diploma or equivalency:1    -0.52981      0.10891  -4.864
## educ33. 12 grades, diploma or equivalency:2    -0.27681      0.16161  -1.713
## educ34. 12 grades, diploma or equivalency, pl:1 -0.77267      0.15382  -5.023
## educ34. 12 grades, diploma or equivalency, pl:2 -0.59665      0.23915  -2.495
## educ35. some college, no degree/ junior/commu:1 -0.90254      0.11128  -8.110
## educ35. some college, no degree/ junior/commu:2 -0.84240      0.16899  -4.985
## educ36. ba level degree:1              -1.10338      0.11704  -9.428
## educ36. ba level degree:2              -1.02332      0.18163  -5.634
## educ37. advanced degrees incl. llb:1         -0.70812      0.12910  -5.485
## educ37. advanced degrees incl. llb:2         -0.97072      0.21290  -4.559
##                                           Pr(>|z|)
## (Intercept):1                          < 2e-16 ***
## (Intercept):2                          0.233432
## ideo3. moderate ('middle of the road'):1      < 2e-16 ***
## ideo3. moderate ('middle of the road'):2      0.120071
## ideo5. conservative:1                    < 2e-16 ***
## ideo5. conservative:2                    < 2e-16 ***
## income2. 17 to 33 percentile:1             0.016762 *
## income2. 17 to 33 percentile:2             0.100354
## income3. 34 to 67 percentile:1             1.96e-08 ***
## income3. 34 to 67 percentile:2             0.001130 **
## income4. 68 to 95 percentile:1             < 2e-16 ***
## income4. 68 to 95 percentile:2             7.67e-07 ***
## income5. 96 to 100 percentile:1             < 2e-16 ***
## income5. 96 to 100 percentile:2             5.27e-06 ***
## educ32. 9-12 grades ('high school'), no diplo:1 0.000332 ***
## educ32. 9-12 grades ('high school'), no diplo:2 0.183085
## educ33. 12 grades, diploma or equivalency:1    1.15e-06 ***
## educ33. 12 grades, diploma or equivalency:2    0.086756 .
## educ34. 12 grades, diploma or equivalency, pl:1 5.09e-07 ***
## educ34. 12 grades, diploma or equivalency, pl:2 0.012599 *
## educ35. some college, no degree/ junior/commu:1 5.04e-16 ***
## educ35. some college, no degree/ junior/commu:2 6.20e-07 ***
## educ36. ba level degree:1                  < 2e-16 ***
## educ36. ba level degree:2                  1.76e-08 ***
## educ37. advanced degrees incl. llb:1         4.13e-08 ***
## educ37. advanced degrees incl. llb:2         5.13e-06 ***
## ---
## 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: 20349.14 on 24726 degrees of freedom
##
## Log-likelihood: -10174.57 on 24726 degrees of freedom
##
## Number of Fisher scoring iterations: 5
##
## No Hauck-Donner effect found in any of the estimates
##

```

```
##  
## Reference group is level 3 of the response  
plot(Reg)
```





2. Explain the results from the fitted model.

*#The predictors seem to be significant.
#The estimates seem to be all negative, thus for a unit increase of Ideo, income,
#education level, the partyid3_b will drop*

3. Use a binned residual plot to assess the fit of the model.

(Optional) Choice models

Using the individual-level survey data from the election example described in Section 10.9 (data available in the folder NES),

1. Fit a logistic regression model for the choice of supporting Democrats or Republicans. Then interpret

the output from this regression in terms of a utility/choice model.

2. Repeat the previous exercise but now with three options: Democrat, no opinion, Republican. That is, fit an ordered logit model and then express it as a utility/choice mode

Contingency table and ordered logit model

In a prospective study of a new living attenuated recombinant vaccine for influenza, patients were randomly allocated to two groups, one of which was given the new vaccine and the other a saline placebo. The responses were titre levels of hemagglutinin inhibiting antibody found in the blood six weeks after vaccination; they were categorized as “small”, “medium” or “large”.

treatment	small	moderate	large	Total
placebo	25	8	5	38
vaccine	6	18	11	35

The cell frequencies in the rows of table are constrained to add to the number of subjects in each treatment group (35 and 38 respectively). We want to know if the pattern of responses is the same for each treatment group.

1. Using a chi-square test and an appropriate log-linear model, test the hypothesis that the distribution of responses is the same for the placebo and vaccine groups.
2. For the model corresponding to the hypothesis of homogeneity of response distributions, calculate the fitted values, the Pearson and deviance residuals, and the goodness of fit statistics X^2 and D . Which of the cells of the table contribute most to X^2 and D ? Explain and interpret these results.
3. Re-analyze these data using ordered logit model (use `polr`) to estimate the cut-points of a latent continuous response variable and to estimate a location shift between the two treatment groups. Sketch a rough diagram to illustrate the model which forms the conceptual base for this analysis.

High School and Beyond

The `hsb` data was collected as a subset of the High School and Beyond study conducted by the National Education Longitudinal Studies program of the National Center for Education Statistics. The variables are gender; race; socioeconomic status; school type; chosen high school program type; scores on reading, writing, math, science, and social studies. We want to determine which factors are related to the choice of the type of program—academic, vocational, or general—that the students pursue in high school. The response is multinomial with three levels.

```
data(hsb)
?hsb
```

1. Fit a trinomial response model with the other relevant variables as predictors (untransformed).
2. For the student with id 99, compute the predicted probabilities of the three possible choices.

Happiness

Data were collected from 39 students in a University of Chicago MBA class and may be found in the dataset `happy`.

```
library(faraway)
data(happy)
```

1. Build a model for the level of happiness as a function of the other variables.
2. Interpret the parameters of your chosen model.

3. Predict the happiness distribution for subject whose parents earn \$30,000 a year, who is lonely, not sexually active and has no job.

Newspaper survey on Vietnam War

A student newspaper conducted a survey of student opinions about the Vietnam War in May 1967. Responses were classified by sex, year in the program and one of four opinions. The survey was voluntary. The data may be found in the dataset `uncviet`. Treat the opinion as the response and the sex and year as predictors. Build a proportional odds model, giving an interpretation to the estimates.

```
data(uncviet)
```

Pneumoconiosis of coal miners

The pneumo data gives the number of coal miners classified by radiological examination into one of three categories of pneumoconiosis and by the number of years spent working at the coal face divided into eight categories.

```
data(pneumo, package = "faraway")
```

1. Treating the pneumoconiosis status as response variable as nominal, build a model for predicting the frequency of the three outcomes in terms of length of service and use it to predict the outcome for a miner with 25 years of service.
2. Repeat the analysis with the pneumoconiosis status being treated as ordinal.
3. Now treat the response variable as hierarchical with top level indicating whether the miner has the disease and the second level indicating, given they have the disease, whether they have a moderate or severe case.
4. Compare the three analyses.