# MA678 homework 05

Multinomial Regression

*Chaoqun Yin*

*October 28, 2018*

## 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 7-point scale) using ideology and demographics with an ordered multinomial logit model.

1. Summarize the parameter estimates numerically and also graphically.

```
x = nes_data_comp$partyid7
nes_data_comp <- nes_data_comp[!is.na(levels(x)[x]),]

nes.order <- polr(factor(partyid7) ~ ideo + age + gender + race + south, Hess = TRUE, data = nes_data_c
summary(nes.order)
```

```
## Call:
## polr(formula = factor(partyid7) ~ ideo + age + gender + race +
##     south, data = nes_data_comp, Hess = TRUE)
##
## Coefficients:
##                         Value Std. Error t value
## ideomoderate          0.95339   0.330586  2.8840
## ideoconservative      1.94046   0.181464 10.6933
## age                  -0.01317   0.004937 -2.6676
## genderfemale         -0.38543   0.155547 -2.4779
## raceblack            -1.79583   0.277242 -6.4775
## raceasian             0.12546   0.544657  0.2303
## racenative american -0.13670   0.368338 -0.3711
## racehispanic         -0.62635   0.297434 -2.1058
## south1                0.21547   0.175126  1.2304
##
## Intercepts:
##                                                          Value   Std. Error
## 1. strong democrat|2. weak democrat                     -1.3844  0.3053
## 2. weak democrat|3. independent-democrat                -0.5338  0.2973
## 3. independent-democrat|4. independent-independent       0.2580  0.2969
## 4. independent-independent|5. independent-republican     0.6528  0.2995
## 5. independent-republican|6. weak republican             1.4496  0.3046
## 6. weak republican|7. strong republican                  2.4281  0.3159
##                                                         t value
## 1. strong democrat|2. weak democrat                     -4.5350
## 2. weak democrat|3. independent-democrat                -1.7958
## 3. independent-democrat|4. independent-independent       0.8689
## 4. independent-independent|5. independent-republican     2.1799
## 5. independent-republican|6. weak republican             4.7597
## 6. weak republican|7. strong republican                  7.6864
##
```

```
## Residual Deviance: 1889.52
## AIC: 1919.52
## (9 observations deleted due to missingness)
```

```
round(summary(nes.order)$coef,2)
```

```
##                                                    Value Std. Error
## ideomoderate                                        0.95       0.33
## ideoconservative                                    1.94       0.18
## age                                                -0.01       0.00
## genderfemale                                       -0.39       0.16
## raceblack                                          -1.80       0.28
## raceasian                                           0.13       0.54
## racenative american                                -0.14       0.37
## racehispanic                                       -0.63       0.30
## south1                                              0.22       0.18
## 1. strong democrat|2. weak democrat                -1.38       0.31
## 2. weak democrat|3. independent-democrat           -0.53       0.30
## 3. independent-democrat|4. independent-independent  0.26       0.30
## 4. independent-independent|5. independent-republican 0.65      0.30
## 5. independent-republican|6. weak republican        1.45       0.30
## 6. weak republican|7. strong republican             2.43       0.32
##                                                    t value
## ideomoderate                                          2.88
## ideoconservative                                     10.69
## age                                                  -2.67
## genderfemale                                         -2.48
## raceblack                                            -6.48
## raceasian                                             0.23
## racenative american                                  -0.37
## racehispanic                                         -2.11
## south1                                                1.23
## 1. strong democrat|2. weak democrat                  -4.54
## 2. weak democrat|3. independent-democrat             -1.80
## 3. independent-democrat|4. independent-independent    0.87
## 4. independent-independent|5. independent-republican  2.18
## 5. independent-republican|6. weak republican          4.76
## 6. weak republican|7. strong republican               7.69
```

2. Explain the results from the fitted model.

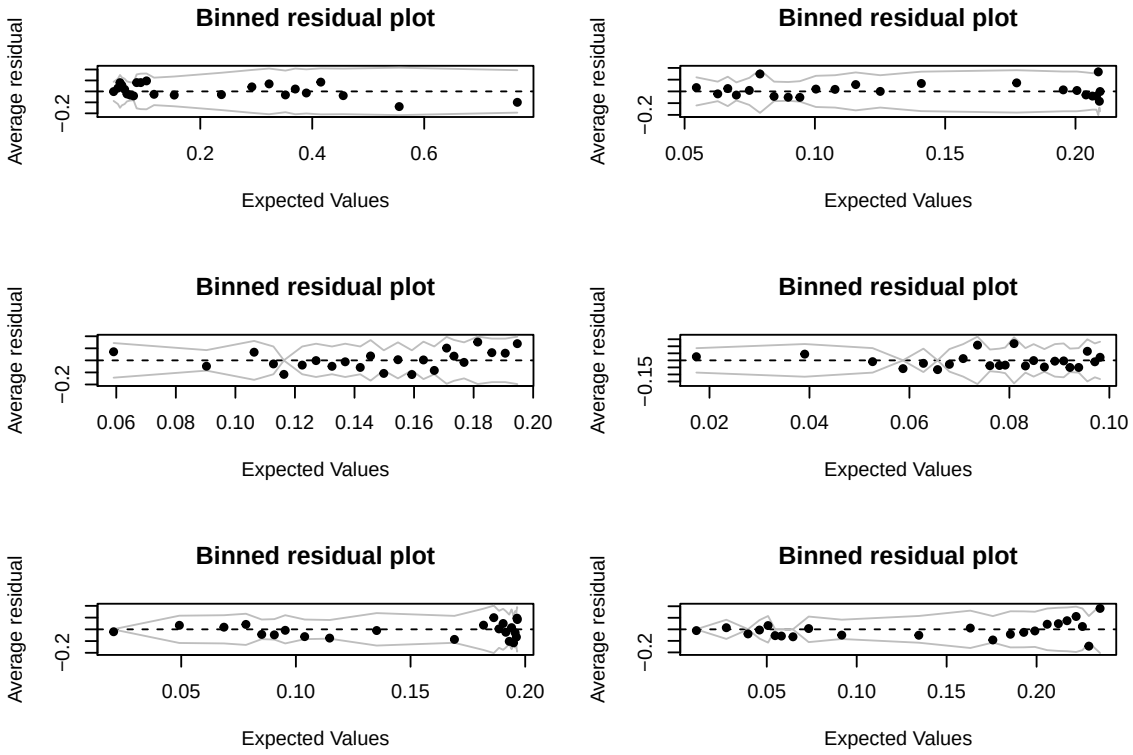
```
confint(nes.order)
```

```
## Waiting for profiling to be done...
```

```
##                          2.5 %       97.5 %
## ideomoderate         0.30274518  1.602885103
## ideoconservative     1.58819102  2.299908250
## age                 -0.02288059 -0.003516047
## genderfemale        -0.69089528 -0.080899207
## raceblack           -2.34876169 -1.259970216
## raceasian           -0.94682823  1.212855171
## racenative american -0.86519074  0.585344750
## racehispanic        -1.21378606 -0.045074705
## south1              -0.12813859  0.558798934
```

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

```
nes <- cbind(partyid7 = nes_data_comp$partyid7, ideo = nes_data_comp$ideo, race = nes_data_comp$race, ag
nes <- data.frame(na.omit(nes))
resid <- model.matrix(~ factor(partyid7) - 1, data = nes) - fitted(nes.order)
par(mfrow = c(3, 2))
for (i in 1:6) {
  binnedplot(fitted(nes.order)[, i], resid[, i], cex.main = 1.3, main = "Binned residual plot")
}
```




# 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).

```
library(nnet)
reg.hs <- multinom(prog ~ read + write + math + science + race,data = hsb,trace = FALSE,HESS = TRUE)
summary(reg.hs)
```

```
## Call:
## multinom(formula = prog ~ read + write + math + science + race,
##     data = hsb, trace = FALSE, HESS = TRUE)
```

```
##
## Coefficients:
##          (Intercept)        read       write       math    science
## general     4.924957 -0.05388450 -0.03946933 -0.1071044 0.09229507
## vocation    8.777829 -0.05594167 -0.06281609 -0.1253231 0.05262485
##           raceasian racehispanic   racewhite
## general  1.11489221  -0.60687283 -0.01313942
## vocation 0.08636574   0.07298783  0.42373684
##
## Std. Errors:
##          (Intercept)       read      write       math    science raceasian
## general     1.528744 0.02853999 0.02864533 0.03391490 0.03053422 0.9950814
## vocation    1.629837 0.03052243 0.02855810 0.03616922 0.03106921 1.3388885
##          racehispanic racewhite
## general     0.8707214 0.6995466
## vocation    0.7864713 0.6836971
##
## Residual Deviance: 332.6696
## AIC: 364.6696
```

2. For the student with id 99, compute the predicted probabilities of the three possible choices.

```
predict(reg.hs,newdata=hsb[hsb$id==99,],type="probs")
```

```
##  academic   general  vocation
## 0.3756043 0.4338602 0.1905356
```

## 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)
?happy
```

1. Build a model for the level of happiness as a function of the other variables.

```
library(nnet)
reg.happy <- polr(factor(happy) ~ money + sex + love + work,data = happy)
summary(reg.happy)
```

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

```
## Call:
## polr(formula = factor(happy) ~ money + sex + love + work, data = happy)
##
## Coefficients:
##          Value Std. Error t value
## money  0.02246    0.01066  2.1064
## sex   -0.47344    0.79498 -0.5955
## love   3.60765    0.80114  4.5031
## work   0.88751    0.40826  2.1739
##
## Intercepts:
```

```
##      Value   Std. Error t value
## 2|3   5.4708  1.9891     2.7504
## 3|4   6.4684  1.9223     3.3650
## 4|5   9.1591  2.1698     4.2212
## 5|6  10.9725  2.3213     4.7268
## 6|7  11.5113  2.3720     4.8530
## 7|8  13.5433  2.6673     5.0776
## 8|9  17.2909  3.1454     5.4971
## 9|10 19.0112  3.3270     5.7142
##
## Residual Deviance: 94.86029
## AIC: 118.8603
```

2. Interpret the parameters of your chosen model.

```
confint(reg.happy)
```

```
## Waiting for profiling to be done...
```

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

```
##              2.5 %     97.5 %
## money  0.002276811 0.04490097
## sex   -2.068912556 1.07918378
## love   2.168908570 5.37172936
## work   0.123787534 1.74622976
```

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

```
predict(reg.happy,newdata = data.frame(love = 1,sex = 0,work = 1,money = 30),type = "probs")
```

```
##            2            3            4            5            6
## 5.749087e-01 2.108348e-01 1.960962e-01 1.515266e-02 1.250656e-03
##            7            8            9           10
## 1.526336e-03 2.252137e-04 4.465166e-06 9.736048e-07
```

## 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)
?uncviet
```

## pneumonoconiosis of coal miners

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

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

```
## Help on topic 'pneumo' was found in the following packages:
##
##   Package               Library
##   VGAM                  /Library/Frameworks/R.framework/Versions/3.4/Resources/library
##   faraway               /Library/Frameworks/R.framework/Versions/3.4/Resources/library
##
##
## Using the first match ...
```

1. Treating the pneumonoconiosis 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.

```
pneumo.reg <- multinom(status ~ year, data = pneumo)
```

```
## # weights:  9 (4 variable)
## initial  value 26.366695
## final  value 26.366695
## converged
```

```
summary(pneumo.reg)
```

```
## Call:
## multinom(formula = status ~ year, data = pneumo)
##
## Coefficients:
##         (Intercept)         year
## normal 2.109424e-15 2.486900e-14
## severe 2.664535e-15 3.552714e-14
##
## Std. Errors:
##        (Intercept)       year
## normal    1.142515 0.03420049
## severe    1.142515 0.03420049
##
## Residual Deviance: 52.73339
## AIC: 60.73339
```

```
predict(pneumo.reg, data.frame (year = 25), type = "probs")
```

```
##      mild    normal    severe
## 0.3333333 0.3333333 0.3333333
```

2. Repeat the analysis with the pneumonoconiosis status being treated as ordinal.

```
pneumo.order <- polr(factor(status) ~ year, data = pneumo, Hess = TRUE)
summary(pneumo.order)
```

```
## Call:
## polr(formula = factor(status) ~ year, data = pneumo, Hess = TRUE)
##
## Coefficients:
##          Value Std. Error   t value
## year 4.341e-11    0.02565 1.692e-09
```

```
##
## Intercepts:
##                Value   Std. Error t value
## mild|normal   -0.6931  0.8838    -0.7842
## normal|severe  0.6931  0.8838     0.7842
##
## Residual Deviance: 52.73339
## AIC: 58.73339
```

```
predict(pneumo.order, data.frame (year = 25), type = "probs")
```

```
##      mild    normal    severe
## 0.3333333 0.3333333 0.3333333
```

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.

```
pneumo$status.h <- ifelse(pneumo$status == "normal", 0, 1)
pneumo.sub <- as.data.frame(cbind(Freq = pneumo$Freq, normal = ifelse(pneumo$status == "normal",1,0), mi

pneumo.hie <- multinom(cbind(normal,mild,severe) ~ year, data = pneumo.sub)
```

```
## # weights:  9 (4 variable)
## initial  value 26.366695
## final  value 26.366695
## converged
```

```
summary(pneumo.hie)
```

```
## Call:
## multinom(formula = cbind(normal, mild, severe) ~ year, data = pneumo.sub)
##
## Coefficients:
##         (Intercept)         year
## mild   1.221245e-15 2.486900e-14
## severe 2.664535e-15 3.552714e-14
##
## Std. Errors:
##        (Intercept)       year
## mild      1.142515 0.03420049
## severe    1.142515 0.03420049
##
## Residual Deviance: 52.73339
## AIC: 60.73339
```

```
predict(pneumo.hie, data.frame(year = 25), type = "probs")
```

```
##    normal      mild    severe
## 0.3333333 0.3333333 0.3333333
```

4. Compare the three analyses. **The results from these three analyses are quite similar, and they are with mild around 0.08-0.10, normal around 0.78-0.83, and severe around 0.08-0.12.**

## (optional) Multinomial choice models:

Pardoe and Simonton (2006) fit a discrete choice model to predict winners of the Academy Awards. Their data are in the folder academy.awards.

| name | description |
|---|---|
| No | unique nominee identifier |
| Year | movie release year (not ceremony year) |
| Comp | identifier for year/category |
| Name | short nominee name |
| PP | best picture indicator |
| DD | best director indicator |
| MM | lead actor indicator |
| FF | lead actress indicator |
| Ch | 1 if win, 2 if lose |
| Movie | short movie name |
| Nom | total oscar nominations |
| Pic | picture nom |
| Dir | director nom |
| Aml | actor male lead nom |
| Afl | actor female lead nom |
| Ams | actor male supporting nom |
| Afs | actor female supporting nom |
| Scr | screenplay nom |
| Cin | cinematography nom |
| Art | art direction nom |
| Cos | costume nom |
| Sco | score nom |
| Son | song nom |
| Edi | editing nom |
| Sou | sound mixing nom |
| For | foreign nom |
| Anf | animated feature nom |
| Eff | sound editing/visual effects nom |
| Mak | makeup nom |
| Dan | dance nom |
| AD | assistant director nom |
| PrNl | previous lead actor nominations |
| PrWl | previous lead actor wins |
| PrNs | previous supporting actor nominations |
| PrWs | previous supporting actor wins |
| PrN | total previous actor/director nominations |
| PrW | total previous actor/director wins |
| Gdr | golden globe drama win |
| Gmc | golden globe musical/comedy win |
| Gd | golden globe director win |
| Gm1 | golden globe male lead actor drama win |
| Gm2 | golden globe male lead actor musical/comedy win |
| Gf1 | golden globe female lead actor drama win |
| Gf2 | golden globe female lead actor musical/comedy win |
| PGA | producer's guild of america win |
| DGA | director's guild of america win |
| SAM | screen actor's guild male win |

| name | description |
|---|---|
| SAF | screen actor's guild female win |
| PN | PP*Nom |
| PD | PP*Dir |
| DN | DD*Nom |
| DP | DD*Pic |
| DPrN | DD*PrN |
| DPrW | DD*PrW |
| MN | MM*Nom |
| MP | MM*Pic |
| MPrN | MM*PrNl |
| MPrW | MM*PrWl |
| FN | FF*Nom |
| FP | FF*Pic |
| FPrN | FF*PrNl |
| FPrW | FF*PrWl |

1. Fit your own model to these data.
2. Display the fitted model on a plot that also shows the data.
3. Make a plot displaying the uncertainty in inferences from the fitted model.