2021-22 MDS-SIM-FINAL

Lídia M.

14/12/2021

Data Description

A study of political ideology and the relationship to sociocultural characterization of U.S. individuals based on an a survey from 944 observations in 10 variables from the 1996 General Social Survey is addressed. Data help quantify well-known relationships between income, age, education, and political affiliation. FPID political affiliation, response variable, is coded into three categories: Democrat, Independent, and Republican. The explanatory variables considered throughout the exercise are: income (factor and covariant), age and education (factor).

Variable and Definitions

- popul population of respondent's location in 1000s of people.
- TVnews days in the past week spent watching news on TV.
- selfLR Left-Right self-placement of respondent: an ordered factor with levels extremely liberal, extLib < liberal, Lib < slightly liberal, sliLib < moderate, Mod < slightly conservative, sliCon < conservative, Con < extremely conservative, extCon.
- ClinLR Left-Right placement of Bill Clinton (same scale as selfLR): an ordered factor with levels extLib < Lib < sliLib < Mod < sliCon < Con < extCon.
- DoleLR Left-Right placement of Bob Dole (same scale as selfLR): an ordered factor with levels extLib < Lib < sliLib < Mod < sliCon < Con < extCon.
- PID Party identification: an ordered factor with levels strong Democrat, strDem < weak Democrat, weakDem < independent Democrat, indDem < independent independentindind < indepedent Republican, indRep < waek Republican, weakRep < strong Republican, strRep.
- age Respondent's age in years
- educ Respondent's education: an ordered factor with levels 8 years or less, MS < high school dropout, HSdrop < high school diploma or GED, HS < some College, Coll < Community or junior College degree, CCdeg < BA degree, BAdeg < postgraduate degree, MAdeg.
- income Respondent's family income: an ordered factor with levels $3Kminus < $3K-$5K < $5K-$7K < $7K-$9K < $9K-$10K < $10K-$11K < $11K-$12K < $12K-$13K < $13K-$14K < $14K-$15K < $15K-$17K < $17K-$20K < $20K-$22K < $22K-$25K < $25K-$30K < $30K-$35K < $35K-$40K < $40K-$45K < $45K-$50K < $50K-$60K < $60K-$75K < $75K-$90K < $90K-$105K < $105Kplus.
- vote Expected vote in 1996 presidential election: a factor with levels Clinton and Dole.
- nincome Salary as a covariant (mid point chosen for each interval).
- FPID Political: Democratic, Independent i Republican – Target – New variable defined for this exam.

Source

Sapiro, Virginia, Steven J. Rosenstone, Donald R. Kinder, Warren E. Miller, and the National Election Studies. AMERICAN NATIONAL ELECTION STUDIES, 1992-1997: COMBINED FILE [Computer file]. 2nd ICPSR version. Ann Arbor, MI: University of Michigan, Center for Political Studies [producer], 1999. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1999.

References

Found at http://www.stat.washington.edu/

Data Preparation and library loading

```
## Loading required package: carData

##
## Attaching package: 'car'

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

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'DescTools'

## The following object is masked from 'package:car':
##
##     Recode

## ResourceSelection 0.3-5   2019-07-22

## Loading required package: data.table

##
## Attaching package: 'data.table'

## The following object is masked from 'package:DescTools':
##
##     %like%

##

## cvAUC version: 1.1.0

## Notice to cvAUC users: Major speed improvements in version 1.1.0

##

##
## Attaching package: 'cvAUC'

## The following object is masked from 'package:DescTools':
##
##     AUC

## null device
##           1

## [1] 944  10

##      popul            TVnews         selfLR       ClinLR       DoleLR
##  Min.   :   0.0   Min.   :0.000   extLib: 16   extLib:109   extLib: 13
```

```
##  1st Qu.:   1.0   1st Qu.:1.000   Lib   :103   Lib   :317   Lib   : 31
##  Median :  22.0   Median :3.000   sliLib:147   sliLib:236   sliLib: 43
##  Mean   : 306.4   Mean   :3.728   Mod   :256   Mod   :160   Mod   : 87
##  3rd Qu.: 110.0   3rd Qu.:7.000   sliCon:170   sliCon: 67   sliCon:195
##  Max.   :7300.0   Max.   :7.000   Con   :218   Con   : 36   Con   :460
##                                   extCon: 34   extCon: 19   extCon:115
##       PID           age            educ           income        vote
##  strDem :200   Min.   :19.00   MS    : 13   $60K-$75K:103   Clinton:551
##  weakDem:180   1st Qu.:34.00   HSdrop: 52   $50K-$60K:100   Dole   :393
##  indDem :108   Median :44.00   HS    :248   $30K-$35K: 70
##  indind : 37   Mean   :47.04   Coll  :187   $25K-$30K: 68
##  indRep : 94   3rd Qu.:58.00   CCdeg : 90   $105Kplus: 68
##  weakRep:150   Max.   :91.00   BAdeg :227   $35K-$40K: 62
##  strRep :175                   MAdeg :127   (Other)  :473

##   popul TVnews selfLR ClinLR DoleLR     PID age  educ   income    vote
## 1     0      7 extCon extLib    Con  strRep  36    HS $3Kminus    Dole
## 2   190      1 sliLib sliLib sliCon weakDem  20  Coll $3Kminus Clinton
## 3    31      7    Lib    Lib    Con weakDem  24 BAdeg $3Kminus Clinton
## 4    83      4 sliLib    Mod sliCon weakDem  28 BAdeg $3Kminus Clinton
## 5   640      7 sliCon    Con    Mod  strDem  68 BAdeg $3Kminus Clinton
## 6   110      3 sliLib    Mod    Con weakDem  21  Coll $3Kminus Clinton

##  [1] "popul"  "TVnews" "selfLR" "ClinLR" "DoleLR" "PID"    "age"    "educ"
##  [9] "income" "vote"

##
##   $3Kminus    $3K-$5K    $5K-$7K    $7K-$9K   $9K-$10K  $10K-$11K  $11K-$12K
##         19         12         17         19         18         13         11
##  $12K-$13K  $13K-$14K  $14K-$15K  $15K-$17K  $17K-$20K  $20K-$22K  $22K-$25K
##         17         10         15         23         35         26         39
##  $25K-$30K  $30K-$35K  $35K-$40K  $40K-$45K  $45K-$50K  $50K-$60K  $60K-$75K
##         68         70         62         48         51        100        103
##  $75K-$90K $90K-$105K  $105Kplus
##         53         47         68

##      popul            TVnews         selfLR       ClinLR       DoleLR
##  Min.   :   0.0   Min.   :0.000   extLib: 16   extLib:109   extLib: 13
##  1st Qu.:   1.0   1st Qu.:1.000   Lib   :103   Lib   :317   Lib   : 31
##  Median :  22.0   Median :3.000   sliLib:147   sliLib:236   sliLib: 43
##  Mean   : 306.4   Mean   :3.728   Mod   :256   Mod   :160   Mod   : 87
##  3rd Qu.: 110.0   3rd Qu.:7.000   sliCon:170   sliCon: 67   sliCon:195
##  Max.   :7300.0   Max.   :7.000   Con   :218   Con   : 36   Con   :460
##                                   extCon: 34   extCon: 19   extCon:115
##       PID           age            educ           income        vote
##  strDem :200   Min.   :19.00   MS    : 13   $60K-$75K:103   Clinton:551
##  weakDem:180   1st Qu.:34.00   HSdrop: 52   $50K-$60K:100   Dole   :393
##  indDem :108   Median :44.00   HS    :248   $30K-$35K: 70
##  indind : 37   Mean   :47.04   Coll  :187   $25K-$30K: 68
##  indRep : 94   3rd Qu.:58.00   CCdeg : 90   $105Kplus: 68
##  weakRep:150   Max.   :91.00   BAdeg :227   $35K-$40K: 62
##  strRep :175                   MAdeg :127   (Other)  :473
##           FPID        nincome
##  Democratic :380   Min.   :  1.50
##  Independent:239   1st Qu.: 23.50
##  Republican :325   Median : 37.50
```

```
##                      Mean   : 46.58
##                      3rd Qu.: 67.50
##                      Max.   :115.00
##
```

The first attempt is a nominal multinomial treatment for FPID target. Answer the questions accurately based on the contents presented in the course and the indicated results of specific models.

Point 1

Determine if the gross effect of the income covariate is statistically significant. Determine if the gross effect of the income covariate is linear on the logodds scale.

Residual deviance for fit.1 model is 1985.424 and 2041.272 for the null model. Deviance difference is 55.85 units assymptotically distributed as Chisq(2), thus H0 stated as 'Both models are equivalent' has a pvalue = 1-pchisq(55.85,2)=7.452927e-13 « 0.05, and H0 can be rejected. Gross effect for income covariate is significant (at any level). A quadratic term of nincome is not needed as seen in the Anova(fit.11) output, it does not add any additional benefit to explain FPID target.

```r
fit.0 <- multinom(FPID ~ 1, data=nes96)
```

```
## # weights:  6 (2 variable)
## initial  value 1037.090001
## final  value 1020.636052
## converged
```

```r
fit.1 <- multinom(FPID~ I(nincome-37.5), data=nes96)
```

```
## # weights:  9 (4 variable)
## initial  value 1037.090001
## final  value 992.712152
## converged
```

```r
anova(fit.0,fit.1)
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: FPID
##                Model Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
## 1                  1      1886   2041.272
## 2 I(nincome - 37.5)      1884   1985.424 1 vs 2     2  55.8478 7.460699e-13
```

```r
Anova(fit.1,test="LR")
```

```
## # weights:  6 (2 variable)
## initial  value 1037.090001
## final  value 1020.636052
## converged

## Analysis of Deviance Table (Type II tests)
##
## Response: FPID
##                     LR Chisq Df Pr(>Chisq)
## I(nincome - 37.5)   55.848  2  7.461e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit.11 <- multinom(FPID~ I(nincome-37.5)+I((nincome-37.5)^2), data=nes96)
```

```
## # weights:  12 (6 variable)
## initial  value 1037.090001
## iter  10 value 991.060037
## iter  10 value 991.060035
## iter  10 value 991.060035
## final  value 991.060035
## converged
```

```r
Anova(fit.11,test="LR")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: FPID
##                        LR Chisq Df Pr(>Chisq)
## I(nincome - 37.5)        32.201  2  1.018e-07 ***
## I((nincome - 37.5)^2)     3.304  2     0.1916
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit.1,fit.11)
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: FPID
##                                      Model Resid. df Resid. Dev   Test    Df
## 1                        I(nincome - 37.5)      1884   1985.424
## 2 I(nincome - 37.5) + I((nincome - 37.5)^2)      1882   1982.120 1 vs 2     2
##   LR stat.   Pr(Chi)
## 1
## 2 3.304235 0.1916437
```

Point 2

Determine if the gross effect of the income factor is significant at the 0.05 significance level.

Residual deviance for fit.2 model (in the addendum) is 1932.417 and 2041.272 for the null model. Deviance difference is 108.86 units assymptotically distributed as Chisq(2X24-2=46), thus H0 stated as 'Both models are equivalent' has a pvalue = 1-pchisq(108.86,46)=5.196456e-07 « 0.05, and H0 can be rejected. Gross effect for income factor is significant (at any level).

```r
fit.2 <- multinom(FPID~ income,data=nes96)
```

```
## # weights:  75 (48 variable)
## initial  value 1037.090001
## iter  10 value 972.715121
## iter  20 value 966.886450
## iter  30 value 966.210960
## iter  40 value 966.208730
## final  value 966.208682
## converged
```

```r
summary(fit.2)
```

```
## Call:
## multinom(formula = FPID ~ income, data = nes96)
```

```
##
## Coefficients:
##             (Intercept) income$3K-$5K income$5K-$7K income$7K-$9K
## Independent   -1.466716     0.9064984     0.3681204     0.4553148
## Republican    -1.466571    -0.4793361     0.8788917     0.4544474
##             income$9K-$10K income$10K-$11K income$11K-$12K income$12K-$13K
## Independent      1.8718341      -0.3250666       0.3680471       1.4674054
## Republican       0.7738364       1.4661561       0.7735543       0.6188759
##             income$13K-$14K income$14K-$15K income$15K-$17K income$17K-$20K
## Independent        1.178962       0.2627070      -0.5480760      -0.3789635
## Republican         1.178981      -0.1430889       0.5501053       1.0871462
##             income$20K-$22K income$22K-$25K income$25K-$30K income$30K-$35K
## Independent       1.2661649       1.0362550       0.3961942       0.1373714
## Republican        0.8606083       0.2631836       0.9556676       1.2359636
##             income$35K-$40K income$40K-$45K income$45K-$50K income$50K-$60K
## Independent        0.847588       0.9559775       0.5499872        1.625774
## Republican         1.299453       1.2434665       1.0199805        1.710187
##             income$60K-$75K income$75K-$90K income$90K-$105K income$105Kplus
## Independent        1.432830        1.312371         1.977888        2.085737
## Republican         1.849594        2.123255         2.405140        2.159749
##
## Std. Errors:
##             (Intercept) income$3K-$5K income$5K-$7K income$7K-$9K
## Independent   0.6405802     0.8962544     0.9245525     0.8667058
## Republican    0.6405426     1.2461623     0.8493499     0.8667823
##             income$9K-$10K income$10K-$11K income$11K-$12K income$12K-$13K
## Independent      0.8295362       1.2557098       1.0378125       0.8342781
## Republican       0.9539943       0.8623415       0.9540776       0.9416579
##             income$13K-$14K income$14K-$15K income$15K-$17K income$17K-$20K
## Independent       0.9968469       0.9185164       0.9883996       0.8923413
## Republican        0.9967954       1.0051785       0.8022773       0.7347443
##             income$20K-$22K income$22K-$25K income$25K-$30K income$30K-$35K
## Independent       0.7825384       0.7329846       0.7226688       0.7422134
## Republican        0.8172287       0.7917744       0.6974855       0.6904656
##             income$35K-$40K income$40K-$45K income$45K-$50K income$50K-$60K
## Independent       0.7212730       0.7373423       0.7418635       0.6886490
## Republican        0.7030009       0.7230465       0.7160972       0.6868802
##             income$60K-$75K income$75K-$90K income$90K-$105K income$105Kplus
## Independent       0.6914910       0.7517358        0.7669051       0.7212728
## Republican        0.6829024       0.7202457        0.7516055       0.7193311
##
## Residual Deviance: 1932.417
## AIC: 2028.417
```

```
anova(fit.0,fit.2,test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: FPID
##    Model Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
## 1      1      1886   2041.272
## 2 income      1840   1932.417 1 vs 2    46 108.8547 5.204762e-07
```

Point 3

Calculate McFadden pseudo coefficient of determination for the model that facilitates the best treatment for the gross income/nincome effect.

AIC(fit.2)=2028.417 and AIC(fit.1)=1993.424, so a numeric treatment including the linear term of nincome covariate gives the best results. According to lecture notes McFadden Pseudo R2 turns out to be 1 - (fit.1deviance/fit.0deviance) = 1-(1985.424/2041.272) = 0.027, thus 2.7% a very low value. There is still plenty of work to be done.

```
AIC(fit.0,fit.1,fit.11,fit.2)
```

```
##        df      AIC
## fit.0   2 2045.272
## fit.1   4 1993.424
## fit.11  6 1994.120
## fit.2  48 2028.417
```

```
# Sheather/McFadden
1 - (fit.1$deviance / fit.0$deviance)
```

```
## [1] 0.02735931
```

```
1 - (fit.2$deviance / fit.0$deviance)
```

```
## [1] 0.05332691
```

Point 4

Once the income is in the model, as a covariant, determine if the net effect of age is statistically significant? Once income and age are in the model, determine if the net effect of education is statistically significant?

According to the provided output, stepwise reduction monitorazed by AIC is given for fit.6 where main effect for educ and interactions to age (linear and quadratic terms) and income are removed, while keeping quadratic term of age that can not retained without the linear term. **(I have not rejected answers indicating this model as the best model since AIC is 1992.88, nevertheless is not convenient)**.

Results for anova(fit.1,fit.5,test="Chisq") test shows the equivalence between numeric and linear income model and the one containing linear and quadratic terms of age (pvalue = 0.23), thus neither age, nor educ are worth to be included.

```
fit.4 <- multinom(FPID~ I(nincome-37.5)+I(age-44), data=nes96)
```

```
## # weights:  12 (6 variable)
## initial  value 1037.090001
## iter  10 value 992.269486
## iter  10 value 992.269484
## iter  10 value 992.269484
## final  value 992.269484
## converged
```

```
fit.5 <- multinom(FPID~ I(nincome-37.5)+I(age-44)+I((age-44)^2), data=nes96)
```

```
## # weights:  15 (8 variable)
## initial  value 1037.090001
## iter  10 value 989.912534
## final  value 989.893728
## converged
```

```r
fit.5p <- multinom(FPID~ I(nincome-37.5)+poly(age,2), data=nes96)
```

```
## # weights:  15 (8 variable)
## initial  value 1037.090001
## iter  10 value 989.965422
## final  value 989.893729
## converged
```

```r
anova(fit.4,fit.5,test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: FPID
##                                               Model Resid. df Resid. Dev   Test
## 1                   I(nincome - 37.5) + I(age - 44)      1882   1984.539
## 2 I(nincome - 37.5) + I(age - 44) + I((age - 44)^2)      1880   1979.787 1 vs 2
##      Df LR stat.    Pr(Chi)
## 1
## 2     2 4.751512 0.09294419
```

```r
anova(fit.1,fit.4,test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: FPID
##                             Model Resid. df Resid. Dev   Test    Df LR stat.
## 1               I(nincome - 37.5)      1884   1985.424
## 2 I(nincome - 37.5) + I(age - 44)      1882   1984.539 1 vs 2     2 0.885336
##     Pr(Chi)
## 1
## 2 0.6423204
```

```r
fit.6p <- multinom(FPID~ (I(nincome-37.5)+poly(age,2))+educ, data=nes96)
```

```
## # weights:  33 (20 variable)
## initial  value 1037.090001
## iter  10 value 989.896378
## iter  20 value 981.755732
## iter  30 value 981.347001
## final  value 981.346988
## converged
```

```r
fit.61 <- multinom(FPID~ (I(nincome-37.5)+poly(age,2))*educ, data=nes96)
```

```
## # weights:  87 (56 variable)
## initial  value 1037.090001
## iter  10 value 1005.987007
## iter  20 value 984.523221
## iter  30 value 973.003769
## iter  40 value 964.442492
## iter  50 value 963.065939
## iter  60 value 962.673812
## iter  70 value 962.369613
## iter  80 value 962.067060
## iter  90 value 961.813037
## iter 100 value 961.681662
## final  value 961.681662
```

```
## stopped after 100 iterations
```

```
fit.6 <- multinom(FPID~ (I(nincome-37.5)+I(age-44)+I((age-44)^2))*educ, data=nes96)
```

```
## # weights:  87 (56 variable)
## initial  value 1037.090001
## iter  10 value 1007.706393
## iter  20 value 995.174772
## iter  30 value 986.782006
## iter  40 value 977.549626
## iter  50 value 962.884569
## iter  60 value 961.407672
## iter  70 value 959.540269
## iter  80 value 959.178630
## final  value 959.170638
## converged
```

```
anova(fit.6p,fit.61,test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: FPID
##                                        Model Resid. df Resid. Dev   Test    Df
## 1 (I(nincome - 37.5) + poly(age, 2)) + educ      1868   1962.694
## 2 (I(nincome - 37.5) + poly(age, 2)) * educ      1832   1923.363 1 vs 2    36
##   LR stat.  Pr(Chi)
## 1
## 2 39.33065 0.323076
```

```
anova(fit.5,fit.6,test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: FPID
##                                                           Model Resid. df
## 1          I(nincome - 37.5) + I(age - 44) + I((age - 44)^2)      1880
## 2 (I(nincome - 37.5) + I(age - 44) + I((age - 44)^2)) * educ      1832
##   Resid. Dev   Test    Df LR stat.    Pr(Chi)
## 1   1979.787
## 2   1918.341 1 vs 2    48 61.44618 0.09201953
```

```
fit.62<-step(fit.6)
```

```
## Start:  AIC=2030.34
## FPID ~ (I(nincome - 37.5) + I(age - 44) + I((age - 44)^2)) *
##     educ
##
## trying - I(nincome - 37.5):educ
## # weights:  69 (44 variable)
## initial  value 1037.090001
## iter  10 value 1012.280762
## iter  20 value 1003.534677
## iter  30 value 982.803143
## iter  40 value 969.029672
## iter  50 value 967.556284
## iter  60 value 966.746676
## iter  70 value 966.344872
```

```
## iter  80 value 964.704873
## final  value 964.698317
## converged
## trying - I(age - 44):educ
## # weights:  69 (44 variable)
## initial  value 1037.090001
## iter  10 value 1008.085509
## iter  20 value 1000.067791
## iter  30 value 988.300605
## iter  40 value 968.050861
## iter  50 value 965.950550
## iter  60 value 965.240113
## iter  70 value 963.985476
## final  value 963.981648
## converged
## trying - I((age - 44)^2):educ
## # weights:  69 (44 variable)
## initial  value 1037.090001
## iter  10 value 999.876521
## iter  20 value 989.988332
## iter  30 value 984.416187
## iter  40 value 967.387610
## iter  50 value 965.499529
## iter  60 value 964.816319
## iter  70 value 962.964116
## iter  80 value 962.952133
## iter  80 value 962.952127
## iter  80 value 962.952127
## final  value 962.952127
## converged
##                          Df      AIC
## - I((age - 44)^2):educ   44 2013.904
## - I(age - 44):educ       44 2015.963
## - I(nincome - 37.5):educ 44 2017.397
## <none>                   56 2030.341
## # weights:  69 (44 variable)
## initial  value 1037.090001
## iter  10 value 999.876521
## iter  20 value 989.988332
## iter  30 value 984.416187
## iter  40 value 967.387610
## iter  50 value 965.499529
## iter  60 value 964.816319
## iter  70 value 962.964116
## iter  80 value 962.952133
## iter  80 value 962.952127
## iter  80 value 962.952127
## final  value 962.952127
## converged
##
## Step:  AIC=2013.9
## FPID ~ I(nincome - 37.5) + I(age - 44) + I((age - 44)^2) + educ +
##     I(nincome - 37.5):educ + I(age - 44):educ
##
```

```
## trying - I((age - 44)^2)
## # weights:  66 (42 variable)
## initial  value 1037.090001
## iter  10 value 1001.189694
## iter  20 value 989.141779
## iter  30 value 985.245105
## iter  40 value 969.711882
## iter  50 value 967.763389
## iter  60 value 967.083886
## iter  70 value 965.533776
## final  value 965.530854
## converged
## trying - I(nincome - 37.5):educ
## # weights:  51 (32 variable)
## initial  value 1037.090001
## iter  10 value 1007.016549
## iter  20 value 986.499009
## iter  30 value 973.920127
## iter  40 value 972.370075
## final  value 972.351424
## converged
## trying - I(age - 44):educ
## # weights:  51 (32 variable)
## initial  value 1037.090001
## iter  10 value 1003.404436
## iter  20 value 988.971511
## iter  30 value 975.884271
## iter  40 value 973.991146
## final  value 973.870987
## converged
##                         Df      AIC
## - I(nincome - 37.5):educ 32 2008.703
## - I(age - 44):educ       32 2011.742
## <none>                   44 2013.904
## - I((age - 44)^2)        42 2015.062
## # weights:  51 (32 variable)
## initial  value 1037.090001
## iter  10 value 1007.016549
## iter  20 value 986.499009
## iter  30 value 973.920127
## iter  40 value 972.370075
## final  value 972.351424
## converged
##
## Step:  AIC=2008.7
## FPID ~ I(nincome - 37.5) + I(age - 44) + I((age - 44)^2) + educ +
##     I(age - 44):educ
##
## trying - I(nincome - 37.5)
## # weights:  48 (30 variable)
## initial  value 1037.090001
## iter  10 value 1013.710886
## iter  20 value 1005.501627
## iter  30 value 998.800990
```

```
## iter  40 value 997.462625
## final  value 997.445974
## converged
## trying - I((age - 44)^2)
## # weights:  48 (30 variable)
## initial  value 1037.090001
## iter  10 value 1000.991813
## iter  20 value 987.961012
## iter  30 value 976.454563
## iter  40 value 974.638647
## final  value 974.632372
## converged
## trying - I(age - 44):educ
## # weights:  33 (20 variable)
## initial  value 1037.090001
## iter  10 value 989.436373
## iter  20 value 981.901784
## final  value 981.346988
## converged
##                     Df      AIC
## - I(age - 44):educ  20 2002.694
## <none>              32 2008.703
## - I((age - 44)^2)   30 2009.265
## - I(nincome - 37.5) 30 2054.892
## # weights:  33 (20 variable)
## initial  value 1037.090001
## iter  10 value 989.436373
## iter  20 value 981.901784
## final  value 981.346988
## converged
##
## Step:  AIC=2002.69
## FPID ~ I(nincome - 37.5) + I(age - 44) + I((age - 44)^2) + educ
##
## trying - I(nincome - 37.5)
## # weights:  30 (18 variable)
## initial  value 1037.090001
## iter  10 value 1012.663626
## iter  20 value 1007.638263
## final  value 1006.839576
## converged
## trying - I(age - 44)
## # weights:  30 (18 variable)
## initial  value 1037.090001
## iter  10 value 988.944213
## iter  20 value 982.349963
## final  value 981.959891
## converged
## trying - I((age - 44)^2)
## # weights:  30 (18 variable)
## initial  value 1037.090001
## iter  10 value 995.486024
## iter  20 value 984.420943
## final  value 984.166274
```

```
## converged
## trying - educ
## # weights:  15 (8 variable)
## initial  value 1037.090001
## iter  10 value 989.912534
## final  value 989.893728
## converged
##                     Df      AIC
## - educ               8 1995.787
## - I(age - 44)       18 1999.920
## <none>              20 2002.694
## - I((age - 44)^2)   18 2004.333
## - I(nincome - 37.5) 18 2049.679
## # weights:  15 (8 variable)
## initial  value 1037.090001
## iter  10 value 989.912534
## final  value 989.893728
## converged
##
## Step:  AIC=1995.79
## FPID ~ I(nincome - 37.5) + I(age - 44) + I((age - 44)^2)
##
## trying - I(nincome - 37.5)
## # weights:  12 (6 variable)
## initial  value 1037.090001
## iter  10 value 1020.409587
## iter  10 value 1020.409587
## iter  10 value 1020.409587
## final  value 1020.409587
## converged
## trying - I(age - 44)
## # weights:  12 (6 variable)
## initial  value 1037.090001
## iter  10 value 990.440541
## iter  10 value 990.440541
## iter  10 value 990.440541
## final  value 990.440541
## converged
## trying - I((age - 44)^2)
## # weights:  12 (6 variable)
## initial  value 1037.090001
## iter  10 value 992.269486
## iter  10 value 992.269484
## iter  10 value 992.269484
## final  value 992.269484
## converged
##                     Df      AIC
## - I(age - 44)        6 1992.881
## <none>               8 1995.787
## - I((age - 44)^2)    6 1996.539
## - I(nincome - 37.5)  6 2052.819
## # weights:  12 (6 variable)
## initial  value 1037.090001
## iter  10 value 990.440541
```

```
## iter  10 value 990.440541
## iter  10 value 990.440541
## final  value 990.440541
## converged
##
## Step:  AIC=1992.88
## FPID ~ I(nincome - 37.5) + I((age - 44)^2)
##
## trying - I(nincome - 37.5)
## # weights:  9 (4 variable)
## initial  value 1037.090001
## final  value 1020.550858
## converged
## trying - I((age - 44)^2)
## # weights:  9 (4 variable)
## initial  value 1037.090001
## final  value 992.712152
## converged
##                     Df      AIC
## <none>               6 1992.881
## - I((age - 44)^2)    4 1993.424
## - I(nincome - 37.5)  4 2049.102
```

Point 5

Determine which of the available proposals is most successful. Determine if the model chosen in the previous point fits well with the data.

The best model obtained so far is the one including the linear term of nincome. Residual deviance is 1985.24 and leaves 2n-p=1888-4 degrees of freedom. Goodness of test fit for the null hypothesis H0 model fits well data has a pvalue of 0.05 (1-pchisq(1985.24, 1888-4)=0.05). Thus H0 is border line to be rejected at 0.05 significance level. Since individual data is provided you can assess goodness of fit by comparing deviance against degrees of freedom and you might conclude that the model is not so bad.

```
anova(fit.1,fit.6,test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: FPID
##                                                        Model Resid. df
## 1                                          I(nincome - 37.5)      1884
## 2 (I(nincome - 37.5) + I(age - 44) + I((age - 44)^2)) * educ      1832
##   Resid. Dev   Test    Df LR stat.   Pr(Chi)
## 1   1985.424
## 2   1918.341 1 vs 2    52 67.08303 0.0777768
```

```
fit.7 <- step(fit.61)
```

```
## Start:  AIC=2035.36
## FPID ~ (I(nincome - 37.5) + poly(age, 2)) * educ
##
## trying - I(nincome - 37.5):educ
## # weights:  69 (44 variable)
## initial  value 1037.090001
## iter  10 value 989.245964
```

```
## iter  20 value 976.235574
## iter  30 value 970.692449
## iter  40 value 969.306489
## iter  50 value 968.833448
## iter  60 value 968.584742
## iter  70 value 968.361148
## iter  80 value 968.261474
## iter  90 value 968.231077
## iter  90 value 968.231076
## final  value 968.231076
## converged
## trying - poly(age, 2):educ
## # weights:  51 (32 variable)
## initial  value 1037.090001
## iter  10 value 1005.988492
## iter  20 value 984.904747
## iter  30 value 976.147360
## iter  40 value 974.003071
## iter  50 value 973.883873
## final  value 973.870987
## converged
##                          Df      AIC
## - poly(age, 2):educ      32 2011.742
## - I(nincome - 37.5):educ 44 2024.462
## <none>                   56 2035.363
## # weights:  51 (32 variable)
## initial  value 1037.090001
## iter  10 value 1005.988492
## iter  20 value 984.904747
## iter  30 value 976.147360
## iter  40 value 974.003071
## iter  50 value 973.883873
## final  value 973.870987
## converged
##
## Step:  AIC=2011.74
## FPID ~ I(nincome - 37.5) + poly(age, 2) + educ + I(nincome -
##     37.5):educ
##
## trying - poly(age, 2)
## # weights:  45 (28 variable)
## initial  value 1037.090001
## iter  10 value 1005.990304
## iter  20 value 985.592467
## iter  30 value 979.725131
## iter  40 value 978.137031
## final  value 978.099234
## converged
## trying - I(nincome - 37.5):educ
## # weights:  33 (20 variable)
## initial  value 1037.090001
## iter  10 value 989.896378
## iter  20 value 981.755732
## iter  30 value 981.347001
```

```
## final  value 981.346988
## converged
##                          Df      AIC
## - I(nincome - 37.5):educ 20 2002.694
## <none>                   32 2011.742
## - poly(age, 2)           28 2012.198
## # weights:  33 (20 variable)
## initial  value 1037.090001
## iter  10 value 989.896378
## iter  20 value 981.755732
## iter  30 value 981.347001
## final  value 981.346988
## converged
##
## Step:  AIC=2002.69
## FPID ~ I(nincome - 37.5) + poly(age, 2) + educ
##
## trying - I(nincome - 37.5)
## # weights:  30 (18 variable)
## initial  value 1037.090001
## iter  10 value 1012.028658
## iter  20 value 1006.869489
## final  value 1006.839580
## converged
## trying - poly(age, 2)
## # weights:  27 (16 variable)
## initial  value 1037.090001
## iter  10 value 990.794848
## iter  20 value 985.873593
## final  value 985.812737
## converged
## trying - educ
## # weights:  15 (8 variable)
## initial  value 1037.090001
## iter  10 value 989.965422
## final  value 989.893729
## converged
##                     Df      AIC
## - educ               8 1995.787
## <none>              20 2002.694
## - poly(age, 2)      16 2003.625
## - I(nincome - 37.5) 18 2049.679
## # weights:  15 (8 variable)
## initial  value 1037.090001
## iter  10 value 989.965422
## final  value 989.893729
## converged
##
## Step:  AIC=1995.79
## FPID ~ I(nincome - 37.5) + poly(age, 2)
##
## trying - I(nincome - 37.5)
## # weights:  12 (6 variable)
## initial  value 1037.090001
```

```
## iter  10 value 1020.410034
## final  value 1020.409587
## converged
## trying - poly(age, 2)
## # weights:  9 (4 variable)
## initial  value 1037.090001
## final  value 992.712152
## converged
##                      Df      AIC
## - poly(age, 2)        4 1993.424
## <none>                8 1995.787
## - I(nincome - 37.5)   6 2052.819
## # weights:  9 (4 variable)
## initial  value 1037.090001
## final  value 992.712152
## converged
##
## Step:  AIC=1993.42
## FPID ~ I(nincome - 37.5)
##
## trying - I(nincome - 37.5)
## # weights:  6 (2 variable)
## initial  value 1037.090001
## final  value 1020.636052
## converged
##                      Df      AIC
## <none>                4 1993.424
## - I(nincome - 37.5)   2 2045.272
```

```r
AIC(fit.1, fit.11, fit.2, fit.4, fit.5, fit.6, fit.7, fit.62)
```

```
##        df      AIC
## fit.1   4 1993.424
## fit.11  6 1994.120
## fit.2  48 2028.417
## fit.4   6 1996.539
## fit.5   8 1995.787
## fit.6  56 2030.341
## fit.7   4 1993.424
## fit.62  6 1992.881
```

```r
fit.1$deviance;fit.1$edf;2*nrow(nes96)-fit.1$edf
```

```
## [1] 1985.424
```

```
## [1] 4
```

```
## [1] 1884
```

```r
1-pchisq(fit.1$deviance, 2*nrow(nes96)-fit.1$edf)
```

```
## [1] 0.05109252
```

Point 6

Interpret the effect of income on logodds and odds scales in the best model available so far.

Best model is fit.1 including linear term of nincome covariate. Interpretation in the logodd scale indicates that each 1000$, one unit increment for nincome around the mean, increases logodds of Independent over Democratic and Republican over Democratic by 0.016 and 0.018 units respectively.

Odds of Independent over Democratic and Republican over Democratic are increased by 1.62% and 1.78%, respectively, for each unit of increment of nincome (1000$) around the mean.

```
summary(fit.1)
```

```
## Call:
## multinom(formula = FPID ~ I(nincome - 37.5), data = nes96)
##
## Coefficients:
##             (Intercept) I(nincome - 37.5)
## Independent  -0.5716824        0.01608684
## Republican   -0.2879428        0.01766452
##
## Std. Errors:
##             (Intercept) I(nincome - 37.5)
## Independent  0.08744296       0.002849736
## Republican   0.08036775       0.002652530
##
## Residual Deviance: 1985.424
## AIC: 1993.424
```

```
coef(fit.1)[,2]
```

```
## Independent  Republican
##  0.01608684  0.01766452
```

```
exp(coef(fit.1)[,2])
```

```
## Independent  Republican
##    1.016217    1.017821
```

```
100*(exp(coef(fit.1)[,2])-1)
```

```
## Independent  Republican
##    1.621693    1.782146
```

Point 7

Calculate estimates for the model parameters in the null multinomial model.

According to the provided output, sample probabilities for Democratic, Independent and Republican are 0.4025424, 0.2531780 and 0.3442797.

Odds for Independent and Republican over Democratic are 0.629 and 0.855 and logodds are -0.464 and -0.156. Output from R confirms these values.

```
table(nes96$FPID)
```

```
##
##  Democratic Independent  Republican
##         380         239         325
```

```
prob<-prop.table(table(nes96$FPID));prob
```

```
##
```

```
##  Democratic Independent  Republican
##   0.4025424   0.2531780   0.3442797
```

```r
oddprob <- prob[2:3]/prob[1];oddprob
```

```
##
## Independent  Republican
##   0.6289474   0.8552632
```

```r
lodd <- log(oddprob); lodd
```

```
##
## Independent  Republican
##  -0.4637077  -0.1563461
```

```r
summary(fit.0)
```

```
## Call:
## multinom(formula = FPID ~ 1, data = nes96)
##
## Coefficients:
##             (Intercept)
## Independent  -0.4636807
## Republican   -0.1563643
##
## Std. Errors:
##             (Intercept)
## Independent  0.08255647
## Republican   0.07555502
##
## Residual Deviance: 2041.272
## AIC: 2045.272
```

Point 8

Calculate the predicted probabilities for the 3 ideologies in 40-year-old women with no education and median income using the best model available so far.

Median nincome is 37.5 so taken the intercept terms is enough to determine the predicted values in the linear predictor scales (logodds Independent or Republican over Democratic) being -0.5716824 and -0.2879428, respectively. Democratic probability has be calculated first as 1/(1+exp(-0.5716824)+exp(-0.2879428))=0.432. Then, Independent probability is Democratic probability by Independent over Democratic odds 0.432x0.565=0.244, and Republican probability is Democratic probability by Republican over Democratic odds 0.432x0.7498045=0.324. Adding up those 3 probabilities 0.432+0.244+0.324 gives 1, as it has to be.

```r
coef(fit.1)
```

```
##             (Intercept) I(nincome - 37.5)
## Independent  -0.5716824        0.01608684
## Republican   -0.2879428        0.01766452
```

```r
predict(fit.1, newdata=data.frame(nincome=37.5),type="probs")
```

```
##  Democratic Independent  Republican
##   0.4320813   0.2439422   0.3239765
```

```r
predict(fit.62, newdata=data.frame(nincome=37.5, age=40),type="probs") # For students using fit.62 as t
```

```
##  Democratic Independent  Republican
##    0.459469    0.237771    0.302760
```

```
logodd <- coef(fit.1)[,1];logodd
```

```
## Independent  Republican
##  -0.5716824  -0.2879428
```

```
odd<-exp(logodd);odd
```

```
## Independent  Republican
##   0.5645748   0.7498045
```

```
pdem <- 1/(1+sum(odd));pdem
```

```
## [1] 0.4320813
```

```
pind <- pdem * odd[1];pind
```

```
## Independent
##   0.2439422
```

```
prep <- pdem * odd[2];prep
```

```
## Republican
##  0.3239765
```

```
pdem+pind+prep
```

```
## Independent
##           1
```

Point 9

Evaluate the predictive power of the model with the chosen nominal response and the improvement over the null model.

This question can be addressed using confusion tables available in the given output. For fit.1 model 284+0+159 observations out of 944 (46.93%) are well-predicted and for fit.0 null model 380 out of 944 (40.25%) are well-predicted. Thus, almost a 6% increment in the predictive power is obtained with respect to the null model.

```
pmprob1 <- predict(fit.1,type="class")
pmprob0 <- predict(fit.0,type="class")
table(pmprob1,nes96$FPID);table(pmprob0,nes96$FPID)
```

```
##
## pmprob1       Democratic Independent Republican
##   Democratic         284         123        166
##   Independent          0           0          0
##   Republican          96         116        159
```

```
##
## pmprob0       Democratic Independent Republican
##   Democratic         380         239        325
##   Independent          0           0          0
##   Republican           0           0          0
```

```
tt<-table(pmprob1,nes96$FPID);tt
```

```
##
```

```
## pmprob1       Democratic Independent Republican
##   Democratic         284         123        166
##   Independent          0           0          0
##   Republican          96         116        159
```

```
tt0<-table(pmprob0,nes96$FPID);tt0
```

```
##
## pmprob0       Democratic Independent Republican
##   Democratic         380         239        325
##   Independent          0           0          0
##   Republican           0           0          0
```

```
100*sum(diag(tt))/sum(tt)
```

```
## [1] 46.92797
```

```
100*sum(diag(tt0))/sum(tt0)
```

```
## [1] 40.25424
```

The second attempt is based on hierarchical logit modelling. A first level defines a binary logit model with a positive response from Other (non-Democratic) and a second level where non-Democratic units are discriminated between Republican (positive response) and Independent. Answer the questions accurately based on the content presented in the course and the indicated results for specific models.

Point 10

Determine whether the gross effect of income is statistically significant? Which would be the best treatment for income in HL1/HL2?

Results can be seen for the first hierarchical level (HL1) using Other as positive outcome and Democratic as negative outcome. Results for mb1.0, mb1.1, mb1.11 and mb1.2 are available (including residual deviance and their degrees of freedom and AIC statistics). Deviance tests can be addressed:

- *Once the linear term of nincome is included in the model, adding the quadratic term is not useful, since H0 'Both models are equivalent' has a pvalue is greater than 0.05 (1271.1-1216.8=0.226) according to their asymptotic distribution Chisq(1): 1-pchisq(0.226,1)=0.635. H0 can not be rejected.*
- *When the linear term of nincome is included in the model, its gross effect is useful, since H0 'Both models are equivalent, null model and mb1.1' has a pvalue is less than 0.05 (1272.6-1217.1=55.491) according to their asymptotic distribution Chisq(1): 1-pchisq(55.491,1)=9.4e-14. H0 can be rejected. Thus nincome linear gross-effect is significant.*
- *Output for mb1.2 including income factor gross-effect is available. The best model between mb1.1 and mb1.2 (not nested) can be addressed by selecting the model with the lowest AIC mm1.1 (AIC(mb1.1)=1221.1 < AIC(mb1.2)=1248.8).*
- *Output for HL2 is also included. Covariate income is not significant, neither linear, no quadratic terms. Nevertheless age covariate (linear and quadratic terms), main effect of educ and some interaction to age provides the best AIC, minimum AIC(after step(mb2.3)=769.9. Using income as a factor output is not available.*

```
nes96$BPID<-nes96$FPID
levels(nes96$BPID)<-c("Democratic","Other","Other")
summary(nes96)
```

```
##       popul              TVnews          selfLR        ClinLR        DoleLR
```

```
##  Min.   :   0.0   Min.   :0.000   extLib: 16   extLib:109   extLib: 13
##  1st Qu.:   1.0   1st Qu.:1.000   Lib   :103   Lib   :317   Lib   : 31
##  Median :  22.0   Median :3.000   sliLib:147   sliLib:236   sliLib: 43
##  Mean   : 306.4   Mean   :3.728   Mod   :256   Mod   :160   Mod   : 87
##  3rd Qu.: 110.0   3rd Qu.:7.000   sliCon:170   sliCon: 67   sliCon:195
##  Max.   :7300.0   Max.   :7.000   Con   :218   Con   : 36   Con   :460
##                                   extCon: 34   extCon: 19   extCon:115
##       PID            age             educ            income        vote
##  strDem :200   Min.   :19.00   MS    : 13   $60K-$75K:103   Clinton:551
##  weakDem:180   1st Qu.:34.00   HSdrop: 52   $50K-$60K:100   Dole   :393
##  indDem :108   Median :44.00   HS    :248   $30K-$35K: 70
##  indind : 37   Mean   :47.04   Coll  :187   $25K-$30K: 68
##  indRep : 94   3rd Qu.:58.00   CCdeg : 90   $105Kplus: 68
##  weakRep:150   Max.   :91.00   BAdeg :227   $35K-$40K: 62
##  strRep :175                   MAdeg :127   (Other)  :473
##           FPID        nincome              BPID
##  Democratic :380   Min.   :  1.50   Democratic:380
##  Independent:239   1st Qu.: 23.50   Other     :564
##  Republican :325   Median : 37.50
##                    Mean   : 46.58
##                    3rd Qu.: 67.50
##                    Max.   :115.00
##
```

```
mb1.0<- glm( BPID~ 1, family=binomial(link="logit") ,data=nes96)
mb1.1<- glm( BPID~ I(nincome-37.5), family=binomial(link="logit") ,data=nes96)
mb1.11<- glm( BPID~ I(nincome-37.5)+I((nincome-37.5)^2), family=binomial(link="logit") ,data=nes96)
mb1.2<- glm( BPID~ income, family=binomial(link="logit") ,data=nes96)
anova(mb1.0,mb1.1,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: BPID ~ 1
## Model 2: BPID ~ I(nincome - 37.5)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       943     1272.6
## 2       942     1217.1  1   55.491 9.389e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(mb1.0,mb1.2,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: BPID ~ 1
## Model 2: BPID ~ income
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       943     1272.6
## 2       920     1200.8 23    71.77 6.462e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(mb1.1,mb1.11,test="Chisq")
```

```
## Analysis of Deviance Table
##
```

```
## Model 1: BPID ~ I(nincome - 37.5)
## Model 2: BPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       942     1217.1
## 2       941     1216.8  1  0.22623   0.6343
```

```
AIC(mb1.0,mb1.1,mb1.11,mb1.2)
```

```
##        df      AIC
## mb1.0   1 1274.567
## mb1.1   2 1221.076
## mb1.11  3 1222.850
## mb1.2  24 1248.797
```

```
summary(mb1.2)
```

```
##
## Call:
## glm(formula = BPID ~ income, family = binomial(link = "logit"),
##     data = nes96)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8182  -1.1557   0.7311   1.0383   1.5183
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.77319    0.49355  -1.567 0.117209
## income$3K-$5K      0.43672    0.76580   0.570 0.568489
## income$5K-$7K      0.65541    0.69260   0.946 0.343999
## income$7K-$9K      0.45474    0.67786   0.671 0.502324
## income$9K-$10K     1.46634    0.70256   2.087 0.036876 *
## income$10K-$11K    0.92734    0.74372   1.247 0.212434
## income$11K-$12K    0.59087    0.78119   0.756 0.449427
## income$12K-$13K    1.12986    0.69746   1.620 0.105237
## income$13K-$14K    1.17865    0.81256   1.451 0.146907
## income$14K-$15K    0.08004    0.73729   0.109 0.913548
## income$15K-$17K    0.14458    0.65974   0.219 0.826534
## income$17K-$20K    0.60134    0.59893   1.004 0.315370
## income$20K-$22K    1.08334    0.63338   1.710 0.087187 .
## income$22K-$25K    0.72190    0.58841   1.227 0.219872
## income$25K-$30K    0.71435    0.54997   1.299 0.193980
## income$30K-$35K    0.83035    0.54843   1.514 0.130016
## income$35K-$40K    1.09861    0.55662   1.974 0.048415 *
## income$40K-$45K    1.10966    0.57385   1.934 0.053148 .
## income$45K-$50K    0.81241    0.56750   1.432 0.152266
## income$50K-$60K    1.66857    0.54052   3.087 0.002022 **
## income$60K-$75K    1.66245    0.53909   3.084 0.002044 **
## income$75K-$90K    1.79769    0.58366   3.080 0.002070 **
## income$90K-$105K   2.21355    0.61727   3.586 0.000336 ***
## income$105Kplus    2.12312    0.57753   3.676 0.000237 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 1272.6  on 943  degrees of freedom
## Residual deviance: 1200.8  on 920  degrees of freedom
## AIC: 1248.8
##
## Number of Fisher Scoring iterations: 4
```

```r
summary(mb1.11)
```

```
##
## Call:
## glm(formula = BPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2),
##     family = binomial(link = "logit"), data = nes96)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8447  -1.2006   0.7411   1.0186   1.3517
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            2.960e-01  8.396e-02   3.525 0.000423 ***
## I(nincome - 37.5)      1.814e-02  3.406e-03   5.326    1e-07 ***
## I((nincome - 37.5)^2) -3.364e-05  7.053e-05  -0.477 0.633381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1272.6  on 943  degrees of freedom
## Residual deviance: 1216.8  on 941  degrees of freedom
## AIC: 1222.8
##
## Number of Fisher Scoring iterations: 4
```

```r
mb2.3<-glm(formula = FPID ~ (I(nincome - 37.5)+I((nincome - 37.5)^2)+I(age - 37.5)+I((age - 37.5)^2))*e
        "Democratic", ])
step(mb2.3)
```

```
## Start:  AIC=793.85
## FPID ~ (I(nincome - 37.5) + I((nincome - 37.5)^2) + I(age - 37.5) +
##     I((age - 37.5)^2)) * educ
##
##                               Df Deviance    AIC
## - I(age - 37.5):educ           5   727.67 785.67
## - I((age - 37.5)^2):educ       5   729.62 787.62
## - I(nincome - 37.5):educ       5   730.45 788.45
## - I((nincome - 37.5)^2):educ   5   730.69 788.69
## <none>                             725.85 793.85
##
## Step:  AIC=785.67
## FPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2) + I(age - 37.5) +
##     I((age - 37.5)^2) + educ + I(nincome - 37.5):educ + I((nincome -
##     37.5)^2):educ + I((age - 37.5)^2):educ
##
##                               Df Deviance    AIC
```

```
## - I((nincome - 37.5)^2):educ  6   732.00 778.00
## - I(nincome - 37.5):educ      6   732.65 778.65
## - I(age - 37.5)               1   727.72 783.72
## - I((age - 37.5)^2):educ      6   738.79 784.79
## <none>                            727.67 785.67
##
## Step:  AIC=778
## FPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2) + I(age - 37.5) +
##     I((age - 37.5)^2) + educ + I(nincome - 37.5):educ + I((age -
##     37.5)^2):educ
##
##                          Df Deviance    AIC
## - I(nincome - 37.5):educ  6   739.31 773.31
## - I(age - 37.5)           1   732.01 776.01
## - I((age - 37.5)^2):educ  6   742.68 776.68
## <none>                        732.00 778.00
## - I((nincome - 37.5)^2)   1   734.45 778.45
##
## Step:  AIC=773.31
## FPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2) + I(age - 37.5) +
##     I((age - 37.5)^2) + educ + I((age - 37.5)^2):educ
##
##                          Df Deviance    AIC
## - I(age - 37.5)           1   739.40 771.40
## - I(nincome - 37.5)       1   741.04 773.04
## <none>                        739.31 773.31
## - I((nincome - 37.5)^2)   1   741.44 773.44
## - I((age - 37.5)^2):educ  6   754.02 776.02
##
## Step:  AIC=771.4
## FPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2) + I((age - 37.5)^2) +
##     educ + I((age - 37.5)^2):educ
##
##                          Df Deviance    AIC
## - I(nincome - 37.5)       1   741.39 771.39
## <none>                        739.40 771.40
## - I((nincome - 37.5)^2)   1   741.61 771.61
## - I((age - 37.5)^2):educ  6   754.09 774.09
##
## Step:  AIC=771.39
## FPID ~ I((nincome - 37.5)^2) + I((age - 37.5)^2) + educ + I((age -
##     37.5)^2):educ
##
##                          Df Deviance    AIC
## - I((nincome - 37.5)^2)   1   741.74 769.74
## <none>                        741.39 771.39
## - I((age - 37.5)^2):educ  6   756.42 774.42
##
## Step:  AIC=769.74
## FPID ~ I((age - 37.5)^2) + educ + I((age - 37.5)^2):educ
##
##                          Df Deviance    AIC
## <none>                        741.74 769.74
## - I((age - 37.5)^2):educ  6   756.67 772.67
```

```
##
## Call:  glm(formula = FPID ~ I((age - 37.5)^2) + educ + I((age - 37.5)^2):educ,
##     family = binomial(link = "logit"), data = nes96[nes96$FPID !=
##         "Democratic", ])
##
## Coefficients:
##                 (Intercept)            I((age - 37.5)^2)
##                    51.18502                     -0.08021
##                  educHSdrop                       educHS
##                   -52.23916                    -51.21422
##                    educColl                   educCCdeg
##                   -50.33742                    -51.17954
##                   educBAdeg                   educMAdeg
##                   -50.66583                    -51.26886
## I((age - 37.5)^2):educHSdrop      I((age - 37.5)^2):educHS
##                     0.08089                      0.08087
##   I((age - 37.5)^2):educColl   I((age - 37.5)^2):educCCdeg
##                     0.07953                      0.08118
##  I((age - 37.5)^2):educBAdeg   I((age - 37.5)^2):educMAdeg
##                     0.08015                      0.08057
##
## Degrees of Freedom: 563 Total (i.e. Null);  550 Residual
## Null Deviance:       768.7
## Residual Deviance: 741.7     AIC: 769.7
```

Point 11

Calculate a 95% confidence interval for the effect of the median range income level factor on the "Democratic" political affiliation odds. Use the mb1.2 model and interpret median range income level in log-odds, odds and probability scales.

Median range income in the model using income factor corresponds to adding dummy estimate for income$35K-$40K being 1.09861 and having an standard error of 0.55662, in the logodds scale can be determined as an increment from the reference group with a 95% confidence interval for median income that ranges from 0.007637089 to 2.19.

Odds for Other (Independent or Republican) over Democratic are multiplied with 95% CI by 1.008 to 8.93 with respect to the reference group (lowest income). Probability of Other is 0.2531780 + 0.3442797 = 0.5974577.

Approximately interpretation increment in the Other probability scale ranges with a 95% CI between 0.007637089x0.5974577x(1-0.5974577)=0.00118 to 2.189587x0.5974577x(1-0.5974577)=0.527 compared to the reference income group (the lowest income).

Instead of interpreting the medium range effect on Other, it is requested for Democratic (the reference level). The logit model taking as positive outcome Democratic would have estimated parameters for dummy variables with reversal sign. So, logodds of Democratic in the medium income range are reduced by 1.099 with respect to the lowest income reference group and odds of Democratic to Other are multiplied by 1/3 or reduced by 66.67% compared to the reference group. And Democratic probability would be approximately reduced by -1.098612x0.5974577x(1-0.5974577)=-0.264 units.

```
ppother <- prop.table(table(nes96$BPID))[2];ppother
```

```
##     Other
## 0.5974576
```

```
coef(mb1.2)[17]
```

```
## income$35K-$40K
##        1.098612
```

```r
exp(coef(mb1.2)[17])
```

```
## income$35K-$40K
##               3
```

```r
coef(mb1.2)[17]*ppother*(1-ppother)
```

```
## income$35K-$40K
##       0.2642185
```

```r
lsup<-coef(mb1.2)[17]+1.96*0.55662;lsup
```

```
## income$35K-$40K
##        2.189587
```

```r
linf<-coef(mb1.2)[17]-1.96*0.55662;linf
```

```
## income$35K-$40K
##     0.007637089
```

```r
exp(c(linf,lsup))
```

```
## income$35K-$40K income$35K-$40K
##        1.007666        8.931528
```

```r
100*(exp(c(linf,lsup))-1)
```

```
## income$35K-$40K income$35K-$40K
##       0.7666326     793.1527998
```

```r
table(nes96$BPID)
```

```
##
## Democratic      Other
##        380        564
```

```r
nes96$BPID <- factor(nes96$BPID, levels = c("Other","Democratic"))
mb1.2i<- glm( BPID~ income, family=binomial(link="logit") ,data=nes96)
summary(mb1.2i)
```

```
##
## Call:
## glm(formula = BPID ~ income, family = binomial(link = "logit"),
##     data = nes96)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5183  -1.0383  -0.7311   1.1557   1.8182
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.77319    0.49355   1.567 0.117209
## income$3K-$5K    -0.43672    0.76580  -0.570 0.568489
## income$5K-$7K    -0.65541    0.69260  -0.946 0.343999
## income$7K-$9K    -0.45474    0.67786  -0.671 0.502324
## income$9K-$10K   -1.46634    0.70256  -2.087 0.036876 *
## income$10K-$11K  -0.92734    0.74372  -1.247 0.212434
```

```
## income$11K-$12K  -0.59087    0.78119  -0.756 0.449427
## income$12K-$13K  -1.12986    0.69746  -1.620 0.105237
## income$13K-$14K  -1.17865    0.81256  -1.451 0.146907
## income$14K-$15K  -0.08004    0.73729  -0.109 0.913548
## income$15K-$17K  -0.14458    0.65974  -0.219 0.826534
## income$17K-$20K  -0.60134    0.59893  -1.004 0.315370
## income$20K-$22K  -1.08334    0.63338  -1.710 0.087187 .
## income$22K-$25K  -0.72190    0.58841  -1.227 0.219872
## income$25K-$30K  -0.71435    0.54997  -1.299 0.193980
## income$30K-$35K  -0.83035    0.54843  -1.514 0.130016
## income$35K-$40K  -1.09861    0.55662  -1.974 0.048415 *
## income$40K-$45K  -1.10966    0.57385  -1.934 0.053148 .
## income$45K-$50K  -0.81241    0.56750  -1.432 0.152266
## income$50K-$60K  -1.66857    0.54052  -3.087 0.002022 **
## income$60K-$75K  -1.66245    0.53909  -3.084 0.002044 **
## income$75K-$90K  -1.79769    0.58366  -3.080 0.002070 **
## income$90K-$105K -2.21355    0.61727  -3.586 0.000336 ***
## income$105Kplus  -2.12312    0.57753  -3.676 0.000237 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1272.6  on 943  degrees of freedom
## Residual deviance: 1200.8  on 920  degrees of freedom
## AIC: 1248.8
##
## Number of Fisher Scoring iterations: 4
```

```
coef(mb1.2i)[17]
```

```
## income$35K-$40K
##       -1.098612
```

```
exp(coef(mb1.2i)[17])
```

```
## income$35K-$40K
##       0.3333333
```

```
coef(mb1.2i)[17]*ppother*(1-ppother)
```

```
## income$35K-$40K
##      -0.2642185
```

```
lsup<-coef(mb1.2i)[17]+1.96*0.55662;lsup
```

```
## income$35K-$40K
##    -0.007637089
```

```
linf<-coef(mb1.2i)[17]-1.96*0.55662;linf
```

```
## income$35K-$40K
##       -2.189587
```

```r
nes96$BPID <- factor(nes96$BPID, levels = c("Democratic","Other"))
```

Point 12

Interpret the effect of the covariate income on the odds scale. Use the mb1.1 and mb2.1 models. Assess whether the proposed mb1.1 and mb2.1 models fit well with data.

(HL1). Each unit increment on numeric income represents an increased of 1.7% in the odds scale of Other to Democratic.

(HL2). Each unit increment on numeric income represents an increased of 0.16% in the odds scale of Republican to Independent conditional to Other. Numeric income is not significant.

Residual deviance and degrees of freedom are given for mb1.1 and mb2.1 models in the output. Goodness of fit test for HL1 does not fulfill asymptotic chi-squared distribution, but deviance(mb1.1)= 1217.1 > df(mb1.1)=942, so missfit is present. Goodness of fit test for HL2 does not fulfill asymptotic chi-squared distribution, but deviance(mb2.1)= 768.3 > df(mb2.1)=562, so missfit is also present.

Hosmer-Lemeshow test is shown for mb1.1 and m2.1 in both cases pvalue « 0.05 indicating that missfit is present in the provided output. Nevertheless, specification of the parameters requires a numeric target, 0 or 1, and once hltest is properly specified pvalues » 0.05 and thus, goodness of fits hold for HL1 and HL2.

```r
mb2.1<-glm(formula = FPID ~ I(nincome - 37.5),
    family = binomial(link = "logit"), data = nes96[nes96$FPID !=
        "Democratic", ])
coef(mb1.1)
```

```
##       (Intercept) I(nincome - 37.5)
##         0.2735652         0.0169970
```

```r
exp(coef(mb1.1))
```

```
##       (Intercept) I(nincome - 37.5)
##          1.314643          1.017142
```

```r
prop.table(table(nes96$BPID))
```

```
##
## Democratic      Other
##  0.4025424  0.5974576
```

```r
coef(mb1.1)*0.5974576*(1-0.5974576)
```

```
##       (Intercept) I(nincome - 37.5)
##       0.065792971       0.004087812
```

```r
coef(mb2.1)
```

```
##       (Intercept) I(nincome - 37.5)
##       0.283465525       0.001595918
```

```r
exp(coef(mb2.1))
```

```
##       (Intercept) I(nincome - 37.5)
##          1.327723          1.001597
```

```r
prop.table(table(nes96$FPID[nes96$BPID!="Democratic"]))
```

```
##
##  Democratic Independent  Republican
```

```
##   0.0000000   0.4237589   0.5762411
```

```r
prop.table(table(nes96$FPID[nes96$BPID!="Democratic"]))*0.5974576
```

```
##
##  Democratic Independent  Republican
##   0.0000000   0.2531780   0.3442796
```

```r
coef(mb2.1)*0.5762411*(1-0.5762411) # Conditional to not being Democratic
```

```
##      (Intercept) I(nincome - 37.5)
##     0.0692186797      0.0003897028
```

```r
mb1.1$deviance;mb1.1$df.residual
```

```
## [1] 1217.076
```

```
## [1] 942
```

```r
1-pchisq(mb1.1$deviance,mb1.1$df.residual)
```

```
## [1] 2.887483e-09
```

```r
mb2.1$deviance;mb2.1$df.residual
```

```
## [1] 768.3443
```

```
## [1] 562
```

```r
1-pchisq(mb2.1$deviance,mb2.1$df.residual)
```

```
## [1] 1.429531e-08
```

```r
library(ResourceSelection)
res.hltest1<-hoslem.test(nes96$BPID, fitted(mb1.1));res.hltest1
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  nes96$BPID, fitted(mb1.1)
## X-squared = 944, df = 8, p-value < 2.2e-16
```

```r
res.hltest2<-hoslem.test(nes96$FPID[nes96$FPID !=
        "Democratic"], fitted(mb2.1));res.hltest2
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  nes96$FPID[nes96$FPID != "Democratic"], fitted(mb2.1)
## X-squared = 564, df = 8, p-value < 2.2e-16
```

```r
# Call using correct arguments
res.hltest1<-hoslem.test(as.numeric(nes96$BPID)-1, fitted(mb1.1));res.hltest1
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  as.numeric(nes96$BPID) - 1, fitted(mb1.1)
## X-squared = 5.403, df = 8, p-value = 0.7138
```

```r
nes96$outhl2 <-ifelse(nes96$FPID=="Independent",0,1)
res.hltest2<-hoslem.test(nes96$outhl2[nes96$FPID !=
```

```r
        "Democratic"], fitted(mb2.1));res.hltest2
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  nes96$outhl2[nes96$FPID != "Democratic"], fitted(mb2.1)
## X-squared = 12.228, df = 8, p-value = 0.1413
```

Point 13

Calculate the predicted probabilities for the 3 ideologies in 40-year-old women with no education and median income. Use the mb1.1 and mb2.1 models

HL1. Medium income is 37.5, thus intercept use is enough. Logodds Other to Democratic is 0.274 and Other probability is exp(coef(mb1.1)[1])/(1+exp(coef(mb1.1)[1]))=0.568. Thus Democratic probability is 1-0.568=0.432.

HL2. Logodds of Republican to Independent for the particular observation is 0.283 and using the response function Republican conditional to Other probability is 0.57 and Independent conditional to Other is 0.43. Thus Republican probability is Other Probability by Republican | Other probability 0.568x0.570=0.324. And Thus Independent probability is Other Probability by Independent | Other probability 0.568x0.430=0.244.

Adding up all probabilities 0.432 + 0.244 + 0.324 is 1, as it should be.

```r
coef(mb1.1)
```

```
##       (Intercept) I(nincome - 37.5)
##         0.2735652         0.0169970
```

```r
exp(coef(mb1.1)[1])
```

```
## (Intercept)
##    1.314643
```

```r
exp(coef(mb1.1)[1])/(1+exp(coef(mb1.1)[1])) # Other
```

```
## (Intercept)
##   0.5679679
```

```r
1-exp(coef(mb1.1)[1])/(1+exp(coef(mb1.1)[1])) # Democratic
```

```
## (Intercept)
##   0.4320321
```

```r
coef(mb2.1)
```

```
##       (Intercept) I(nincome - 37.5)
##       0.283465525       0.001595918
```

```r
exp(coef(mb2.1)[1])
```

```
## (Intercept)
##    1.327723
```

```r
exp(coef(mb2.1)[1])/(1+exp(coef(mb2.1)[1])) # Republican | Other
```

```
## (Intercept)
##   0.5703956
```

```r
1-exp(coef(mb2.1)[1])/(1+exp(coef(mb2.1)[1])) # Independent | Other
```

```
## (Intercept)
##   0.4296044
```

```r
(exp(coef(mb2.1)[1])/(1+exp(coef(mb2.1)[1])))*exp(coef(mb1.1)[1])/(1+exp(coef(mb1.1)[1])) # Republican
```

```
## (Intercept)
##   0.3239664
```

```r
(1-exp(coef(mb2.1)[1])/(1+exp(coef(mb2.1)[1])))*exp(coef(mb1.1)[1])/(1+exp(coef(mb1.1)[1])) # Independe
```

```
## (Intercept)
##   0.2440015
```

Point 14

Determine the estimated parameters for HL1 and HL2 probit null models.

Democratic probability in the sample is 0.4025 thus probability of Other is 0.5975 and Republican | Other is 0.5762. Probit transformation for these probabilities are qnorm(c(0.5975,0.5762))-> 0.247 and 0.192, respectively, thus constant estimates for null logit models HL1 and HL2. Results from R for comparison purposes can be seen below.

```r
prop.table(table(nes96$BPID))
```

```
##
## Democratic      Other
##  0.4025424  0.5974576
```

```r
prop.table(table(nes96$FPID[nes96$BPID!="Democratic"]))
```

```
##
##  Democratic Independent  Republican
##   0.0000000   0.4237589   0.5762411
```

```r
pother0 <-prop.table(table(nes96$BPID))[2]
pdem0 <-prop.table(table(nes96$BPID))[1];pdem0
```

```
## Democratic
##  0.4025424
```

```r
prep0 <-prop.table(table(nes96$FPID[nes96$BPID!="Democratic"]))[3];prep0
```

```
## Republican
##  0.5762411
```

```r
pind0 <-prop.table(table(nes96$FPID[nes96$BPID!="Democratic"]))[2];pind0
```

```
## Independent
##   0.4237589
```

```r
# log(pother0/(1-pother0))
# log(prep0/(1-prep0))
qnorm(pother0)
```

```
##     Other
## 0.2467719
```

```r
qnorm(prep0)
```

```
## Republican
##  0.1922866
```

```r
# Not given in the output, just for validation purposes
mb1.0 <- glm(formula = BPID ~ 1,
    family = binomial(link = "probit"), data = nes96)
mb2.0 <- glm(formula = FPID ~ 1,
    family = binomial(link = "probit"), data = nes96[nes96$FPID !=
        "Democratic", ])

summary(mb1.0)
```

```
##
## Call:
## glm(formula = BPID ~ 1, family = binomial(link = "probit"), data = nes96)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.349  -1.349   1.015   1.015   1.015
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.24677    0.04125   5.983 2.19e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1272.6  on 943  degrees of freedom
## Residual deviance: 1272.6  on 943  degrees of freedom
## AIC: 1274.6
##
## Number of Fisher Scoring iterations: 4
```

```r
summary(mb2.0)
```

```
##
## Call:
## glm(formula = FPID ~ 1, family = binomial(link = "probit"), data = nes96[nes96$FPID !=
##     "Democratic", ])
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
##  -1.31   -1.31    1.05    1.05    1.05
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.19229    0.05313   3.619 0.000296 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 768.71  on 563  degrees of freedom
## Residual deviance: 768.71  on 563  degrees of freedom
## AIC: 770.71
##
## Number of Fisher Scoring iterations: 3
```

The third attempt is the proportional odds modelling, estimated from the point of view of latent variable in R. Answer the questions accurately based on the contents presented in the subject and the indicated results of specific models.

Point 15

Determine which of the available proposals for proportional odds modelling is most successful. Assess whether the proposed model fits well with the data.

Linear term of nincome gross effect (model om1) is signficant according to deviance test anova(om0, om1, test="Chisq") « 0.05 given in the output. Adding quadratic term of nincome to om1 is shown on om11 and testing its net-effect shows a pvalue of 0.06 (border line). Nevertheless, it can be seen in the output AIC(om1) > AIC(om11). So, linear and quadratic terms are retained under the numeric treatment.

Income factor model om2 shows AIC(om2)=2030 > AIC(om11)=2000. Then, om11 seems to be the best modelling proposal available so far for ordinal treatment.

Goodness of fit for any om1, om11 and om2 show residual deviances being greater than residual degrees of freedom. So, missfit is present.

```r
om1<-polr(formula = FPID ~ I(nincome - 37.5), data = nes96)
om11<-polr(formula = FPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2),data = nes96)
om2<-polr(formula = FPID ~ income,data = nes96)
summary(om2)
```

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

## Call:
## polr(formula = FPID ~ income, data = nes96)
##
## Coefficients:
##                     Value Std. Error t value
## income$3K-$5K    0.23858     0.7360 0.32418
## income$5K-$7K    0.73827     0.6779 1.08904
## income$7K-$9K    0.44389     0.6629 0.66957
## income$9K-$10K   0.98221     0.6313 1.55577
## income$10K-$11K  1.28607     0.7386 1.74128
## income$11K-$12K  0.65075     0.7635 0.85227
## income$12K-$13K  0.82984     0.6487 1.27926
## income$13K-$14K  1.07597     0.7553 1.42458
## income$14K-$15K  0.03002     0.7243 0.04144
## income$15K-$17K  0.28309     0.6557 0.43174
## income$17K-$20K  0.85357     0.5942 1.43662
## income$20K-$22K  0.89491     0.6021 1.48637
## income$22K-$25K  0.54347     0.5703 0.95303
## income$25K-$30K  0.80645     0.5409 1.49091
## income$30K-$35K  1.04404     0.5408 1.93064
## income$35K-$40K  1.14254     0.5426 2.10554
## income$40K-$45K  1.10373     0.5565 1.98343
## income$45K-$50K  0.88261     0.5561 1.58704
```

```
## income$50K-$60K  1.44925     0.5198 2.78828
## income$60K-$75K  1.57614     0.5205 3.02789
## income$75K-$90K  1.83508     0.5538 3.31360
## income$90K-$105K 1.90705     0.5574 3.42108
## income$105Kplus  1.69399     0.5337 3.17433
##
## Intercepts:
##                        Value  Std. Error t value
## Democratic|Independent 0.7380 0.4873     1.5145
## Independent|Republican 1.8345 0.4901     3.7432
##
## Residual Deviance: 1980.027
## AIC: 2030.027
```

```
AIC(om1,om11, om2)
```

```
##      df      AIC
## om1   3 2001.363
## om11  4 1999.956
## om2  25 2030.027
```

```
om1$deviance;om1$edf;2*nrow(nes96)-om1$edf
```

```
## [1] 1995.363
## [1] 3
## [1] 1885
```

```
om11$deviance;om11$edf;2*nrow(nes96)-om11$edf
```

```
## [1] 1991.956
## [1] 4
## [1] 1884
```

```
om2$deviance;om2$edf;2*nrow(nes96)-om2$edf
```

```
## [1] 1980.027
## [1] 25
## [1] 1863
```

Point 16

Interpret the effect of the income covariate on the scale of proportional odds for om1 model.

Latent variable paradigm applies for polr() method in MASS library, so proportional odds estimates for explanatory variables need to reverse signs. Odds for Democratic to Rest or Independent to Republican are decreased by 100x(1-exp(-coef(om1)[1]))= 1.3% for each unit increment in covariate nincome.

Since om1 summary was not available in the outputs it was indicated during the exam to interpret the lineal term of om11.

```
summary(om1)
```

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

```
## Call:
## polr(formula = FPID ~ I(nincome - 37.5), data = nes96)
##
## Coefficients:
##                     Value Std. Error t value
## I(nincome - 37.5) 0.01312    0.00197   6.659
##
## Intercepts:
##                         Value   Std. Error t value
## Democratic|Independent -0.2829  0.0693    -4.0826
## Independent|Republican  0.7996  0.0737    10.8517
##
## Residual Deviance: 1995.363
## AIC: 2001.363
```

```
-coef(om1)[1]
```

```
## I(nincome - 37.5)
##       -0.01311985
```

```
100*(1-exp(-coef(om1)[1]))
```

```
## I(nincome - 37.5)
##          1.303416
```

```
summary(om11)
```

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

```
## Call:
## polr(formula = FPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2),
##     data = nes96)
##
## Coefficients:
##                           Value Std. Error t value
## I(nincome - 37.5)      0.017573  0.0042284  4.1559
## I((nincome - 37.5)^2) -0.000108  0.0001286 -0.8401
##
## Intercepts:
##                         Value   Std. Error t value
## Democratic|Independent -0.3557  0.1034    -3.4386
## Independent|Republican  0.7299  0.1034     7.0561
##
## Residual Deviance: 1991.956
## AIC: 1999.956
```

```
-coef(om11)[1]
```

```
## I(nincome - 37.5)
##       -0.01757277
```

```
100*(1-exp(-coef(om11)[1]))
```

```
## I(nincome - 37.5)
##          1.741927
```

Point 17

Calculate the predicted probabilities for the 3 ideologies in 40-year-old women with no education and median income for om11 model

Again, only nincome covariate is included in om11 model and it is centered on median income (37.5), so only intercept estimates have be taken into account. Cumulative log-odds are -0.3557 and 0.73, thus logit response function has to be applied to these values to get cumulative probabilities for Democratic and Democratic+Independent, being 0.412 and 0.675. Thus, predicted probability based on om11 model for the requested observation would be Democratic probability 0.412, Independent probability 0.675-0.412=0.263 and Republican probability 1-0.675=0.325. These values are validated using generic predict method in R.

```
predict(om11,newdata=data.frame(nincome=37.5),type="probs")
```

```
##  Democratic Independent  Republican
##   0.4120029   0.2627722   0.3252250
```

```
summary(om11)
```

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

## Call:
## polr(formula = FPID ~ I(nincome - 37.5) + I((nincome - 37.5)^2),
##     data = nes96)
##
## Coefficients:
##                            Value Std. Error t value
## I(nincome - 37.5)       0.017573  0.0042284  4.1559
## I((nincome - 37.5)^2) -0.000108  0.0001286 -0.8401
##
## Intercepts:
##                         Value   Std. Error t value
## Democratic|Independent -0.3557  0.1034    -3.4386
## Independent|Republican  0.7299  0.1034     7.0561
##
## Residual Deviance: 1991.956
## AIC: 1999.956
```

```
clogodds <- om11$zeta; clogodds
```

```
## Democratic|Independent Independent|Republican
##             -0.3556916              0.7298623
```

```
expclogodds <- exp( clogodds ); expclogodds
```

```
## Democratic|Independent Independent|Republican
##              0.7006886              2.0747948
```

```
probs<-expclogodds/(1+expclogodds);probs
```

```
## Democratic|Independent Independent|Republican
##              0.4120029              0.6747750
```

```
probdem <- probs[1]
probind <- probs[2]-probs[1]
probrep <- 1-probs[2]
probdem;probind;probrep
```

```
## Democratic|Independent
##               0.4120029
```

```
## Independent|Republican
##               0.2627722
```

```
## Independent|Republican
##                0.325225
```

Point 18

Assess the predictive power of the chosen model in the output using ordinal response and the improvement over the null model.

Confusion tables for a non-specified model and null proportional odds models are given in the output. Accuracy for the null model is 40.25% and 46.93% for the alternative model. So it is very similar to the values obtained before.

```r
poprobx <- predict(om2,type="class")
om0<-polr(formula = FPID ~ 1, data = nes96)
poprob0 <- predict(om0,type="class")
table(poprobx,nes96$FPID);table(poprob0,nes96$FPID)
```

```
##
## poprobx       Democratic Independent Republican
##   Democratic         284         123        166
##   Independent          0           0          0
##   Republican          96         116        159
```

```
##
## poprob0       Democratic Independent Republican
##   Democratic         380         239        325
##   Independent          0           0          0
##   Republican           0           0          0
```

```r
100*sum(diag(table(poprobx,nes96$FPID)))/sum(table(poprobx,nes96$FPID))
```

```
## [1] 46.92797
```

```r
100*sum(diag(table(poprob0,nes96$FPID)))/sum(table(poprob0,nes96$FPID))
```

```
## [1] 40.25424
```

Point 19

Determine the estimated zeta (cut-off) points for the null model in the ordinal proposal when logit and probit link functions are stated

Cumulative probabilities for Democratic and Democratic+Independent in the sample are 0.403 and 0.656.

Logit transformation has to be applied to cumulative sample probabilities to get intercepts estimates for null model, being -0.395 and 0.644, respectively.

Probit transformation has to be applied to cumulative sample probabilities to get intercepts estimates for null model, being -0.247 and 0.401, respectively.

Results are consistent with estimates provided by R.

```
prop.table(table(nes96$FPID))
```

```
##
##  Democratic Independent  Republican
##   0.4025424   0.2531780   0.3442797
```

```
cprop <-c(prop.table(table(nes96$FPID))[1],prop.table(table(nes96$FPID))[1]+prop.table(table(nes96$FPID
```

```
## Democratic Democratic
##  0.4025424  0.6557203
```

```
names(cprop)<-c("Democratic", "Independent")
codds <- cprop/(1-cprop);codds
```

```
##  Democratic Independent
##   0.6737589   1.9046154
```

```
logcodds <- log(codds);logcodds
```

```
##  Democratic Independent
##  -0.3948830   0.6442801
```

```
probitcum <- qnorm(cprop); probitcum
```

```
##  Democratic Independent
##  -0.2467719   0.4008109
```

```
om0<-polr(formula = FPID ~ 1, data = nes96)
om0$zeta
```

```
## Democratic|Independent Independent|Republican
##             -0.3948650              0.6442962
```

```
om0p<-polr(formula = FPID ~ 1, data = nes96, method="probit")
om0p$zeta
```

```
## Democratic|Independent Independent|Republican
##             -0.2467718              0.4008095
```

Point 20

Which of the 3 proposals, nominal multinomial, hierarchical logit or proportional logodds, is the most satisfactory?

Best models for nominal, HL and ordinal treatment are fit.1, mb1.1-mb2.1 and om11. AIC statistics can be found in the output to be 1993.424, 1993.42 and 1999.956, respectively. So, Hierarchical logit treatment or Multinomial treatments give the lowest AIC statistic. Additionally, if available a discussion about accuracy would be useful to take the final decision. Taking a look to accuracy and/or F1-Score is convenient, but you do not have available all the needed data. HL has a priori advantage, it pays more attention to Independent units, that fail to be predicted with Multinomial and Ordinal treatments, again you can not use quantitatively this argument.

```
AIC(fit.1)
```

```
## [1] 1993.424
```

```
AIC(mb1.1)+AIC(mb2.1)
```

```
## [1] 1993.42
```

```r
AIC(om11)
```

```
## [1] 1999.956
```

```r
100*sum(diag(table(poprobx,nes96$FPID)))/sum(table(poprobx,nes96$FPID))
```

```
## [1] 46.92797
```