Homework4

Yufei Yin

Part A.

The table below relates political ideology to political party affiliation. Political ideology has a five-point ordinal scaled ranging from very liberal to very conservative: X =political party; Y =political ideology; Z =gender. The data are used by Q14 of Exercises for Chapter 3 in Homework 4.

1.

Calculate the sample marginal odds ratios of Democratic vs Republican between females and males, and intepret it.

```
set1 <- read.csv('pol_ideol_data.csv')
c.table <- xtabs(formula = count ~ party + gender , data = set1)
c.table
```

```
##      gender
## party   F   M
##     D 264 164
##     R 219 188
```

```
round((264*188)/(219*164),2)
```

```
## [1] 1.38
```

The odds that female in Democrat party is estimated to be 1.38 times the odds that female in Republican party.

2.

Calculate the sample conditional odds ratios of Democratic vs Republican between females and males with political ideology fixed at each of the five levels, and intepret them.

very liberal

```
c.table <- xtabs(formula = count ~ party + gender , data = set1[set1$ideol == "VL",])
c.table
```

```
##      gender
## party  F  M
##     D 44 36
##     R 18 12
```

```
round((44*12)/(18*36),2)
```

```
## [1] 0.81
```

Given political ideology is very liberal, The odds that female in Democrat party is estimated to be 0.81 times the odds that female in Republican party

slightly liberal

```
c.table <- xtabs(formula = count ~ party + gender , data = set1[set1$ideol == "SL",])
c.table
```

```
##      gender
## party  F  M
##     D 47 34
##     R 28 18
```

```
round((47*18)/(28*34),2)
```

```
## [1] 0.89
```

Given political ideology is slightly liberal, The odds that female in Democrat party is estimated to be 0.89 times the odds that female in Republican party

moderate

```
c.table <- xtabs(formula = count ~ party + gender , data = set1[set1$ideol == "M",])
c.table
```

```
##      gender
## party   F   M
##     D 118  53
##     R  86  62
```

```
round((118*62)/(86*53),2)
```

```
## [1] 1.61
```

Given political ideology is Moderate, The odds that female in Democrat party is estimated to be 1.61 times the odds that female in Republican party

slightly conservative

```
c.table <- xtabs(formula = count ~ party + gender , data = set1[set1$ideol == "SC",])
c.table
```

```
##      gender
## party  F  M
##     D 23 18
##     R 39 45
```

```
round((23*45)/(39*18),2)
```

```
## [1] 1.47
```

Given political ideology is slightly conservative, The odds that female in Democrat party is estimated to be 1.47 times the odds that female in Republican party

very conservative

```
c.table <- xtabs(formula = count ~ party + gender , data = set1[set1$ideol == "VC",])
c.table
```

```
##      gender
## party  F  M
##     D 32 23
##     R 48 51
```

```
round((32*51)/(48*23),2)
```

```
## [1] 1.48
```

Given political ideology is very conservative, The odds that female in Democrat party is estimated to be 1.48 times the odds that female in Republican party

3.

Conduct the Cohran-Mantel-Haenszel test for X - Z given Y , the indepdence of gender and political party affiliation conditional on political ideology. Comment on test outcome.

```
c.table <- xtabs(formula = count ~ party + gender + ideol , data = set1)
c.table
```

```
## , , ideol = M
##
##      gender
## party   F   M
##     D 118  53
##     R  86  62
##
## , , ideol = SC
##
##      gender
## party   F   M
##     D  23  18
##     R  39  45
##
## , , ideol = SL
##
##      gender
## party   F   M
##     D  47  34
##     R  28  18
##
## , , ideol = VC
##
##      gender
## party   F   M
##     D  32  23
##     R  48  51
##
## , , ideol = VL
##
##      gender
## party   F   M
##     D  44  36
##     R  18  12
```

```
mantelhaen.test(c.table, correct=F)
```

```
##
##  Mantel-Haenszel chi-squared test without continuity correction
##
## data:  c.table
## Mantel-Haenszel X-squared = 3.5124, df = 1, p-value = 0.06091
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.987835 1.754986
## sample estimates:
## common odds ratio
##          1.316676
```

Testing for Conditional Independence

p-value = 0.06091 larger than 0.05 we reject the null hypothesis political party and gender are independent given political ideology.

4.

Conduct the Breslow-Day test for the homogeneity of the odds ratios X - Z conditional on Y . Comment on the test outcome.

```r
breslowday.test(c.table)
```

```
## Breslow and Day test (with Tarone correction):
## Breslow-Day X-squared         = 3.235357
## Breslow-Day-Tarone X-squared  = 3.23528
##
## Test for test of a common OR: p-value =  0.5192516
```

Testing for Homogeneity of Conditional OR

p-value = 0.5192516 larger than 0.05 we reject the null hypothesis conditional OR of political party and gender does not depend on political ideology.

Q14

An example from Section 4.2.5 examines data from the 1991 U.S. General Social Survey that cross-classifies people according to

- Political ideology: Very liberal (VL), Slightly liberal (SL),Moderate (M), Slightly conservative (SC), and Very conservative (VC)
- Political party: Democrat (D) or Republican (R)
- Gender: Female (F) or Male (M).

Consider political ideology to be an ordinal response variable, and political party and gender to be explanatory variables. The data are available in the file pol_ideol_data.csv.

(a)

Use the factor() function with the ideology variable to make sure that R places the levels of the ideology variable in the correct order.

```
set1 <- read.csv('pol_ideol_data.csv')
set1$ideol <- factor(set1$ideol, levels = c("VL", "SL", "M", "SC", "VC"))
```

(b)

Because the two explanatory variables are categorical, we can view the entire data set in a three-dimensional contingency table structure. The xtabs() and ftable() functions are useful for this purpose. Below is how these functions can be used after the data file has been read into R (the data frame set1 contains the original data):

```
c.table <- xtabs(formula = count ~ party + ideol + gender , data = set1)
c.table
```

```
## , , gender = F
##
##      ideol
## party  VL  SL   M  SC  VC
##     D  44  47 118  23  32
##     R  18  28  86  39  48
##
## , , gender = M
##
##      ideol
## party  VL  SL   M  SC  VC
##     D  36  34  53  18  23
##     R  12  18  62  45  51
```

```
flat.c.table <- ftable(x = c.table , row.vars = c("gender", "party") , col.vars = "ideol")
flat.c.table
```

```
##              ideol  VL  SL   M  SC  VC
## gender party
## F      D            44  47 118  23  32
##        R            18  28  86  39  48
## M      D            36  34  53  18  23
##        R            12  18  62  45  51
```

(c)

Using multinomial and proportional odds regression models that include party, gender, and their interaction, complete the following:

i.

Estimate the models and perform LRTs to test the importance of each explanatory variable

multinomial

```
library(nnet)
mod.fit.nom <- multinom(ideol ~ party*gender,
                        weights = count,
                        data=set1)
```

```
## # weights:  25 (16 variable)
## initial  value 1343.880657
## iter  10 value 1231.244704
## iter  20 value 1229.548447
## final  value 1229.543342
## converged
```

```
round(coefficients(mod.fit.nom),4)
```

```
##    (Intercept) partyR genderM partyR:genderM
## SL      0.0660 0.3759 -0.1232         0.0868
## M       0.9865 0.5775 -0.5998         0.6780
## SC     -0.6487 1.4219 -0.0444         0.5929
## VC     -0.3184 1.2992 -0.1297         0.5958
```

$log(\hat{\pi}_{SL}/\hat{\pi}_{VL}) = 0.0660 + 0.3759R - 0.1232M + 0.0868R \times M$

$log(\hat{\pi}_{M}/\hat{\pi}_{VL}) = 0.9865 + 0.5775R - 0.5998M + 0.6780R \times M$

$log(\hat{\pi}_{SC}/\hat{\pi}_{VL}) = -0.6487 + 1.4219R - 0.0444M + 0.5929R \times M$

$log(\hat{\pi}_{VC}/\hat{\pi}_{VL}) = -0.3184 + 1.2992R - 0.1297M + 0.5958R \times M$

```
library(car)
Anova(mod.fit.nom)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: ideol
##              LR Chisq Df Pr(>Chisq)
## party          60.555  4  2.218e-12 ***
## gender          8.965  4    0.06198 .
## party:gender    3.245  4    0.51763
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction term is not important given other explanatory variables in the model.

proportional

```
library(MASS)
mod.fit.po<-polr(ideol ~ party*gender,
```

```r
                 weights = count,
                 method = "logistic",
                 data=set1)
summary(mod.fit.po)
```

```
## Call:
## polr(formula = ideol ~ party * gender, data = set1, weights = count,
##     method = "logistic")
##
## Coefficients:
##                  Value Std. Error t value
## partyR          0.7562     0.1659  4.5593
## genderM        -0.1431     0.1820 -0.7861
## partyR:genderM  0.5091     0.2550  1.9965
##
## Intercepts:
##       Value    Std. Error t value
## VL|SL  -1.5521   0.1332   -11.6560
## SL|M   -0.5550   0.1157    -4.7965
## M|SC    1.1647   0.1226     9.5009
## SC|VC   2.0012   0.1364    14.6666
##
## Residual Deviance: 2470.15
## AIC: 2484.15
```

$logti(\hat{P}(Y \leq j)) = \hat{\beta}_{j0} - 0.7562R + 0.1431M - 0.5091R \times M$

$\hat{\beta}_{10} = -1.5521$

$\hat{\beta}_{20} = -0.5550$

$\hat{\beta}_{30} = 1.1647$

$\hat{\beta}_{40} = 2.0012$

```r
Anova(mod.fit.po)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: ideol
##              LR Chisq Df Pr(>Chisq)
## party          56.847  1  4.711e-14 ***
## gender          0.843  1    0.35864
## party:gender    3.992  1    0.04571 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

gender is not important given other explanatory variables in the model.

ii.

Compute the estimated probabilities for each ideology level given all possible combinations of the party and gender levels.

multinomial

```
data <- as.data.frame(flat.c.table)
pi.hat <- predict(object = mod.fit.nom, newdata = data[1:4,], type = "probs")
nom.pi.hat <- data.frame(gender = data[1:4,1],
                         party = data[1:4,2],
                         round(pi.hat,4))
nom.pi.hat
```

```
##   gender party     VL     SL      M     SC     VC
## 1      F     D 0.1667 0.1780 0.4470 0.0871 0.1212
## 2      M     D 0.2195 0.2073 0.3232 0.1098 0.1402
## 3      F     R 0.0822 0.1279 0.3927 0.1781 0.2192
## 4      M     R 0.0638 0.0957 0.3298 0.2394 0.2713
```

proportional

```
pi.hat <- predict(object = mod.fit.po, newdata = data[1:4,], type = "probs")
po.pi.hat <- data.frame(gender = data[1:4,1],
                        party = data[1:4,2],
                        round(pi.hat,4))
po.pi.hat
```

```
##   gender party     VL     SL      M     SC     VC
## 1      F     D 0.1748 0.1899 0.3975 0.1187 0.1191
## 2      M     D 0.1964 0.2021 0.3887 0.1080 0.1049
## 3      F     R 0.0904 0.1218 0.3884 0.1757 0.2236
## 4      M     R 0.0645 0.0930 0.3531 0.1960 0.2934
```

iii.

Construct a contingency table with estimated counts from the model. These estimated counts are found by taking the estimated probability for each ideology level multiplied by their corresponding number of observations for a party and gender combination. For example, there are 264 observations for gender = "F" and party = "D". Because the multinomial regression model results in $\hat{\pi}_{VL} = 0.1667$, this model's estimated count is $0.1667 \times 264 = 44$.

multinomial

```
sum <- aggregate(set1$count, by=list(gender=set1$gender, party = set1$party), FUN=sum)
FD <- sum$x[1] * as.numeric(nom.pi.hat[1,3:7])
MD <- sum$x[2] * as.numeric(nom.pi.hat[2,3:7])
FR <- sum$x[3] * as.numeric(nom.pi.hat[3,3:7])
MR <- sum$x[4] * as.numeric(nom.pi.hat[4,3:7])
estimated.count <- c(FD,FR,MD,MR)
set2 <- cbind(set1,estimated.count)
c.table <- xtabs(formula = estimated.count ~ party + ideol + gender , data = set2)
ftable(x = c.table , row.vars = c("gender", "party") , col.vars = "ideol")
```

```
##              ideol       VL       SL        M       SC       VC
## gender party
## F      D            44.0088  46.9920 118.0080  22.9944  31.9968
##        R            18.0018  28.0101  86.0013  39.0039  48.0048
## M      D            35.9980  33.9972  53.0048  18.0072  22.9928
##        R            11.9944  17.9916  62.0024  45.0072  51.0044
```

proportional

```
FD <- sum$x[1] * as.numeric(po.pi.hat[1,3:7])
MD <- sum$x[2] * as.numeric(po.pi.hat[2,3:7])
FR <- sum$x[3] * as.numeric(po.pi.hat[3,3:7])
MR <- sum$x[4] * as.numeric(po.pi.hat[4,3:7])
estimated.count <- c(FD,FR,MD,MR)
set3 <- cbind(set1,estimated.count)
c.table <- xtabs(formula = estimated.count ~ party + ideol + gender , data = set3)
(count <- ftable(x = c.table , row.vars = c("gender", "party") , col.vars = "ideol"))
```

```
##              ideol       VL       SL        M       SC       VC
## gender party
## F      D            46.1472  50.1336 104.9400  31.3368  31.4424
##        R            19.7976  26.6742  85.0596  38.4783  48.9684
## M      D            32.2096  33.1444  63.7468  17.7120  17.2036
##        R            12.1260  17.4840  66.3828  36.8480  55.1592
```

iv.

Are the estimated counts the same as the observed? Explain

For multinomial model, it is the same as the observed, because the model is saturated.

For proportional model, it is not the same as the observed, because the model is not saturated.

v.

Use odds ratios computed from the estimated models to help understand relationships between the explanatory variables and the response.

multinomial

```
round(exp(coefficients(mod.fit.nom)[,-1]), 2)
```

```
##    partyR genderM partyR:genderM
## SL   1.46    0.88           1.09
## M    1.78    0.55           1.97
## SC   4.15    0.96           1.81
## VC   3.67    0.88           1.81
```

```
conf.beta <- confint(object = mod.fit.nom , level = 0.95)
round (exp( conf.beta [2:4 , , ]) ,2)
```

```
## , , SL
##
##                2.5 % 97.5 %
## partyR          0.71   2.99
## genderM         0.47   1.65
## partyR:genderM  0.35   3.37
##
## , , M
##
##                2.5 % 97.5 %
## partyR          0.96   3.29
## genderM         0.32   0.95
## partyR:genderM  0.75   5.19
##
## , , SC
##
##                2.5 % 97.5 %
## partyR          1.95   8.80
## genderM         0.45   2.04
## partyR:genderM  0.58   5.64
##
## , , VC
##
##                2.5 % 97.5 %
## partyR          1.81   7.44
## genderM         0.44   1.76
## partyR:genderM  0.62   5.35
```

proportional

```
round(exp(-coefficients(mod.fit.po)), 2)
```

```
##         partyR        genderM partyR:genderM
##           0.47           1.15           0.60
```

```
conf.beta <- confint(object = mod.fit.po , level = 0.95)
ci <- exp (- conf.beta )
round ( data.frame(low = ci[,2] , up = ci[ ,1]) , 2)
```

```
##                  low   up
## partyR         0.34 0.65
## genderM        0.81 1.65
## partyR:genderM 0.36 0.99
```

(d)

Compare the results for the two models. Discuss which model may be more appropriate for this setting.

The proportional odds regression model is more appropriate, because we have ordinal response vraible. we use factor() function in question 1 to make sure that R places the levels of the ideology variable in the correct order.

15

Continuing Exercise 14, consider again the counts that are found using the estimated proportional odds model. Find the correct combination of counts from this table that result in estimated odds that are the same as the estimated odds found directly from the estimated model $\exp(\hat{\beta}_{j0} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1 x_2)$, where x_1 and x_2 are the appropriate indicator variables representing party and gender, respectively.

```
count
```

```
##               ideol       VL       SL         M       SC       VC
## gender party
## F      D            46.1472  50.1336 104.9400  31.3368  31.4424
##        R            19.7976  26.6742  85.0596  38.4783  48.9684
## M      D            32.2096  33.1444  63.7468  17.7120  17.2036
##        R            12.1260  17.4840  66.3828  36.8480  55.1592
```

```
data <- as.data.frame(count)
(FD.data <- data[data$gender == "F"& data$party == "D",])
```

```
##    gender party ideol     Freq
## 1       F     D    VL  46.1472
## 5       F     D    SL  50.1336
## 9       F     D     M 104.9400
## 13      F     D    SC  31.3368
## 17      F     D    VC  31.4424
```

```
FD.SUM <- sum(FD.data$Freq)
```

given female, democrate, the estimated odds VL|SL

```
vl <- FD.data$Freq[1]
vl/(FD.SUM-vl)
```

```
## [1] 0.2118274
```

```
summary(mod.fit.po)
```

```
##
## Re-fitting to get Hessian
## Call:
## polr(formula = ideol ~ party * gender, data = set1, weights = count,
##     method = "logistic")
##
## Coefficients:
##                  Value Std. Error t value
## partyR          0.7562     0.1659  4.5593
## genderM        -0.1431     0.1820 -0.7861
## partyR:genderM  0.5091     0.2550  1.9965
##
## Intercepts:
##       Value    Std. Error t value
## VL|SL  -1.5521   0.1332   -11.6560
## SL|M   -0.5550   0.1157    -4.7965
## M|SC    1.1647   0.1226     9.5009
## SC|VC   2.0012   0.1364    14.6666
##
## Residual Deviance: 2470.15
## AIC: 2484.15
```

recall:

$logti(\hat{P}(Y \leq j)) = \hat{\beta}_{j0} - 0.7562R + 0.1431M - 0.5091R \times M$

$\hat{\beta}_{10} = -1.5521$

$\hat{\beta}_{20} = -0.5550$

$\hat{\beta}_{30} = 1.1647$

$\hat{\beta}_{40} = 2.0012$

```
exp(-1.5521)
```

```
## [1] 0.2118027
```

given female, democrate, the estimated odds of SL|M

```
sl<- sum(FD.data$Freq[1:2])
sl/(FD.SUM-sl)
```

```
## [1] 0.5740595
```

```
exp(-0.5550)
```

```
## [1] 0.5740723
```

given female, democrate, the estimated odds of M|SC

```
m <- sum(FD.data$Freq[1:3])
m/(FD.SUM-m)
```

```
## [1] 3.205214
```

```
exp(1.1647)
```

```
## [1] 3.204961
```

given female, democrate, the etimated odds of SC|VC

```
sc.lower <- sum(FD.data$Freq[1:4])
(sc.lower/FD.SUM)/((FD.SUM-sc.lower)/FD.SUM)
```

```
## [1] 7.396306
```

```
exp(2.0012)
```

```
## [1] 7.397928
```

According the calculation above, the correct combination of counts from this table that result in estimated odds that are the same as the estimated odds found directly from the estimated model