P8131 HW4

Shihui Zhu, sz3029

1. Summarize the data using appropriate tables of percentages to show the pairwise associations between the levels of satisfaction and 1) contact with other residents and 2) type of housing. Comment on patterns in the associations.

```
value = c(65, 130, 67, 34, 141, 130, 54, 76, 48, 47, 116, 105, 100, 111, 62, 100, 191, 104)
data1 <- tibble(
  Contact = c(rep("Low", 3), rep("High", 3), rep("Low", 3), rep("High", 3), rep("Low", 3), rep("High", 3
  Satisfaction = c(rep("Low", 6), rep("Medium", 6), rep("High", 6)),
  ApartmentType = c("Tower Block", "Apartment", "House", "Tower Block", "Apartment", "House",
                    "Tower Block", "Apartment", "House", "Tower Block", "Apartment", "House",
                    "Tower Block", "Apartment", "House", "Tower Block", "Apartment", "House")
)
data1 = data1[rep(seq_len(nrow(data1)), value),]
```

Produce Summary Table:

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1681
##
##
##                     | data1$ApartmentType
## data1$Satisfaction |   Apartment |       House | Tower Block |   Row Total |
## -------------------|-------------|-------------|-------------|-------------|
##               High |         302 |         166 |         200 |         668 |
##                    |       0.013 |       7.437 |      10.600 |             |
##                    |       0.452 |       0.249 |       0.299 |       0.397 |
##                    |       0.395 |       0.322 |       0.500 |             |
##                    |       0.180 |       0.099 |       0.119 |             |
## -------------------|-------------|-------------|-------------|-------------|
##                Low |         271 |         197 |          99 |         567 |
##                    |       0.652 |       3.027 |       9.563 |             |
```

```
##                     |       0.478 |       0.347 |       0.175 |       0.337 |
##                     |       0.354 |       0.382 |       0.247 |             |
##                     |       0.161 |       0.117 |       0.059 |             |
## --------------------|-------------|-------------|-------------|-------------|
##              Medium |         192 |         153 |         101 |         446 |
##                     |       0.593 |       1.892 |       0.248 |             |
##                     |       0.430 |       0.343 |       0.226 |       0.265 |
##                     |       0.251 |       0.297 |       0.253 |             |
##                     |       0.114 |       0.091 |       0.060 |             |
## --------------------|-------------|-------------|-------------|-------------|
##       Column Total |         765 |         516 |         400 |        1681 |
##                     |       0.455 |       0.307 |       0.238 |             |
## --------------------|-------------|-------------|-------------|-------------|
##
##


##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1681
##
##
##                     | data1$Contact
## data1$Satisfaction |      High |       Low | Row Total |
## --------------------|-----------|-----------|-----------|
##                High |       395 |       273 |       668 |
##                     |     0.278 |     0.377 |           |
##                     |     0.591 |     0.409 |     0.397 |
##                     |     0.408 |     0.383 |           |
##                     |     0.235 |     0.162 |           |
## --------------------|-----------|-----------|-----------|
##                 Low |       305 |       262 |       567 |
##                     |     1.416 |     1.923 |           |
##                     |     0.538 |     0.462 |     0.337 |
##                     |     0.315 |     0.367 |           |
##                     |     0.181 |     0.156 |           |
## --------------------|-----------|-----------|-----------|
##              Medium |       268 |       178 |       446 |
##                     |     0.486 |     0.660 |           |
##                     |     0.601 |     0.399 |     0.265 |
##                     |     0.277 |     0.250 |           |
##                     |     0.159 |     0.106 |           |
## --------------------|-----------|-----------|-----------|
##       Column Total |       968 |       713 |      1681 |
##                     |     0.576 |     0.424 |           |
```

```
## -------------------|-----------|-----------|-----------|
##
##
```

From the table we see that

2. Nominal logistic regression model

Use nominal logistic regression model for the associations between response variable, the levels of satisfaction, and the other two variables.

Obtain a model that summarizes the patterns in the data.

Construct a nominal logistic regression model

```
data1.mult <- multinom(cbind(Sat.Low, Sat.Medium, Sat.High) ~ ApartmentType + Contact, data = data1.sat)
```

```
## # weights:  15 (8 variable)
## initial  value 1846.767257
## iter  10 value 1803.046285
## final  value 1802.740161
## converged
```

```
summary(data1.mult)
```

```
## Call:
## multinom(formula = cbind(Sat.Low, Sat.Medium, Sat.High) ~ ApartmentType +
##     Contact, data = data1.sat)
##
## Coefficients:
##            (Intercept) ApartmentTypeHouse ApartmentTypeTower Block ContactLow
## Sat.Medium  -0.2180364         0.06967922                0.4067631 -0.2959832
## Sat.High     0.2474047        -0.30402275                0.6415948 -0.3282264
##
## Std. Errors:
##            (Intercept) ApartmentTypeHouse ApartmentTypeTower Block ContactLow
## Sat.Medium  0.10930968          0.1437749                0.1713009  0.1301046
## Sat.High    0.09783068          0.1351693                0.1500774  0.1181870
##
## Residual Deviance: 3605.48
## AIC: 3621.48
```

Goodness of fit:

```
# goodness of fit
pihat=predict(data1.mult,type='probs')
m=rowSums(data1.sat[,3:5])
res.pearson=(data1.sat[,3:5]-pihat*m)/sqrt(pihat*m);res.pearson # pearson residuals
```

```
##      Sat.Low  Sat.Medium   Sat.High
## 1  0.6462082  0.01458006 -0.4986448
## 2  0.3770510  0.08967620 -0.4648120
```

```
## 3 -1.0575683 -0.12653898  1.4047956
## 4 -0.8014220 -0.01559243  0.5248140
## 5 -0.3508834 -0.07196683  0.3670803
## 6  0.8402535  0.08670506 -0.9471979
```

```
G.stat=sum(res.pearson^2) # Generalized Pearson Chisq Stat
G.stat
```

```
## [1] 6.932341
```

```
pval=1-pchisq(G.stat,df=(6-4)*(3-1))
pval# fit is good
```

```
## [1] 0.1395072
```

```
# deviance
D.stat = sum(2*data1.sat[,3:5]*log(data1.sat[,3:5]/(pihat*m)))
D.stat
```

```
## [1] 6.893028
```

The Generalized Pearson Chisq Statistics is 6.932341. The p-value is $0.1395072 > 0.05$, so we can reject the null hypothesis and the model fit is good. The Deviance is 6.893028.

```
pi_low <- 1/(1+sum(exp(coef(data1.mult))))
pi_medium <- sum(exp(coef(data1.mult)[c(1,3,5,7)]))/(1+sum(exp(coef(data1.mult))))
pi_high <- sum(exp(coef(data1.mult)[c(2,4,6,8)]))/(1+sum(exp(coef(data1.mult))))
```

Describe your findings (the pattern in the associations, odds ratios with 95% confidence intervals, goodness-of-fit). (Hint: use dummy variable for house types.) Is there interaction of contact level by house type?

```
invfisher.mult <- vcov(data1.mult) # inverse of fisher information matrix
CI.logit.medium = coef(data1.mult)[c(1, 3, 5, 7)] + kronecker(t(c(0,qnorm(0.025),-qnorm(0.025))),
                                                  t(t(sqrt(diag(invfisher.mult)[1:4]))))
CI.logit.high = coef(data1.mult)[c(2, 4, 6, 8)] + kronecker(t(c(0,qnorm(0.025),-qnorm(0.025))),
                                                  t(t(sqrt(diag(invfisher.mult)[5:8]))))

out.pi_low <- cbind(pi_low,
                    1/(1+sum(exp(CI.logit.medium[2:4,2]) + exp(CI.logit.high[2:4,2]))),
                    1/(1+sum(exp(CI.logit.medium[2:4,3]) + exp(CI.logit.high[2:4,3])))
                    )
out.pi_medium <- cbind(pi_medium,
                (sum(exp(CI.logit.medium[2:4,2])))/(1+sum(exp(CI.logit.medium[2:4,2]) + exp(CI.logit.high
                (sum(exp(CI.logit.medium[2:4,3])))/(1+sum(exp(CI.logit.medium[2:4,3]) + exp(CI.logit.high
                )

out.pi_high <- cbind(pi_high,
                (sum(exp(CI.logit.high[2:4,2])))/(1+sum(exp(CI.logit.medium[2:4,2]) + exp(CI.logit.high[2
                (sum(exp(CI.logit.high[2:4,3])))/(1+sum(exp(CI.logit.medium[2:4,3]) + exp(CI.logit.high[2
                )
out <- rbind(out.pi_low, out.pi_medium, out.pi_high)
colnames(out)=c('Estimate of Odds Ratio','95% CI lower','95% CI upper')
out %>% knitr::kable(digits = 3)
```

Estimate of Odds Ratio	95% CI lower	95% CI upper
0.102	0.166	0.101
0.422	0.409	0.453
0.475	0.425	0.446

```
data1.grouped$Satisfaction = factor(data1.grouped$Satisfaction, levels = c("Low", "Medium", "High"), or
data1.grouped$Contact = factor(data1.grouped$Contact, levels = c("Low", "High"), ordered=T)
data1.grouped$ApartmentType = as.factor(data1.grouped$ApartmentType)

data1.polr=polr(Satisfaction ~ ApartmentType + Contact, data = data1.grouped, weights = Value)
summary(data1.polr)
```

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

## Call:
## polr(formula = Satisfaction ~ ApartmentType + Contact, data = data1.grouped,
##     weights = Value)
##
## Coefficients:
##                            Value Std. Error t value
## ApartmentTypeHouse       -0.2353     0.1052  -2.236
## ApartmentTypeTower Block  0.5010     0.1168   4.291
## Contact.L                 0.1785     0.0658   2.713
##
## Intercepts:
##             Value   Std. Error t value
## Low|Medium  -0.6226  0.0721    -8.6347
## Medium|High  0.4899  0.0714     6.8575
##
## Residual Deviance: 3610.286
## AIC: 3620.286
```

The model tells us

Goodness of fit and discrepency:

```
pihat=predict(data1.polr,data1.sat,type='p')
m=rowSums(data1.sat[,3:5])
res.pearson=(data1.sat[,3:5]-pihat*m)/sqrt(pihat*m);res.pearson # pearson residuals
```

```
##      Sat.Low Sat.Medium    Sat.High
## 1  0.7794163 -0.3696759 -0.31516502
## 2  0.9176717 -1.0671397 -0.01522921
## 3 -1.1408504  0.1397991  1.24412460
## 4 -0.9946605  0.4549798  0.33539244
## 5 -0.2370110 -0.4051905  0.53781037
## 6  0.2742957  1.3678375 -1.47778315
```

```r
G=sum(res.pearson^2)
G
```

```
## [1] 11.64205
```

```r
numsamp=(3-1)*6 # degree of freedom for grouped data
numparam=2+3 # total num of param
pval=1-pchisq(G ,df=numsamp-numparam)
pval # fits well
```

```
## [1] 0.112962
```

The p-value is $0.112962 > 0.05$, so the model fits the data well.

The largest error are