P8131 Spring 2022 Homework #4 Solution

The data in the table below are from an investigation into satisfaction with housing conditions in Copenhagen (Madsen 1971). Residents in selected areas living in rented homes built between 1960 and 1968 were questioned about their satisfaction and the degree of contact with other residents.

	Low satisfaction	Medium satisfaction	High satisfaction
Contact with others:	Low High	Low High	Low High
Tower block	65 34	54 47	100 100
Apartment	130 141	76 116	111 191
House	67 130	48 105	62 104

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

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

3. As the response has ordinal categories, fit proportional odds model to the data that include the same variables as used in the nominal logistic model obtained in (ii). What does the fitted model tell?

4. Calculate Pearson residuals from the proportional odds model for ordinal response to find where the largest discrepancies are between the observed frequencies and expected frequencies estimated from the model.

Solution:

1. The summary tables for pair-wise associations between the three variables.

   1) Percentages of levels of satisfaction with housing conditions by contact with other residents.

Contact/Satisfaction	Low	Medium	High
High (n=968)	31.51%	27.69%	40.81%
Low (n=713)	36.75%	24.96%	38.29%

The distribution of levels of satisfaction with housing conditions seemed not very different by levels of contact with other residents.
(Test null hypothesis of no association between contact with other residents and level of satisfaction with housing conditions: Chi-Square test statistic $T_S = 5.14, df = 2, p = 0.0765$; Cannot reject null hypothesis. Therefore, levels of satisfaction and contact with other residents are independent (α=0.05).)

2) Percentages of various level of satisfaction with housing conditions by type of housing.

Housing/Satisfaction	Low	Medium	High
Apartment (n=765)	35.42%	25.10%	39.48%
House (n=516)	38.18%	29.65%	32.17%
Tower block (n=400)	24.75%	25.25%	50.00%

Tower block residents seemed to have higher percent of high level satisfaction and lower percent of low level satisfaction than that of those living in apartment or those living in house.
(Test null hypothesis of no association between house type and level of satisfaction with housing conditions: Chi-Square test $T_S = 34.02, df = 4, p < .0001$; reject null hypothesis. Therefore, level of satisfaction with housing conditions depends on type of housing.)

2. 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. Describe your findings (the pattern in the associations, odds ratios with 95% confidence intervals, goodness-of-fit). (Hint: use dummy variable for house types)
(optional; required for PhD) Is there interaction of contact level by house type?
Fit nominal models to the data with dummy variables: $C = 1$ for high and 0 for low contact. $TypeH = 1$ for house, 0 else; $TypeT = 1$ for tower block, 0 else. Response categories of satisfaction with housing conditions: 1 for low level, 2 for medium level, 3 for high level.
Model 1 (m_1):

$$\log\left(\frac{\pi_j(X)}{\pi_1(X)}\right) = \alpha_j + \beta_{j1}C + \beta_{j2}TypeH + \beta_{j3}TypeT, j = 2, 3$$

Model 2 (m_2):

$$\log\left(\frac{\pi_j(X)}{\pi_1(X)}\right) = \alpha_j + \beta_{j1}C + \beta_{j2}TypeH + \beta_{j3}TypeT + \beta_{j4}C * TypeH + \beta_{j5}C * TypeT, j = 2, 3$$

The fitted model m_1 suggested that house type and contact level were related to satisfaction level. The null hypothesis that house type was unrelated to satisfaction

level, i.e., $H_0 : \beta_{j2} = \beta_{j3} = 0; j = 2, 3$, was rejected, as Wald test $T_S = 36.87, df = 4, p < .0001$. And the null hypothesis that contact level was unrelated to satisfaction level, i.e., $H_0 : \beta_{j1} = 0; j = 2, 3$, was rejected as Wald test $T_S = 8.86, df = 2, p = 0.0119$. Model m_1 had Pearson residuals ranged between -1.057 and 1.405, resulting in a Pearson Chi- square statistic of 6.93 and deviance of 6.89 with $df = 4, p = 0.14$. Model m_1 seemed to fit data well and could be used to describe the data. Specifically, for each house type, high level of contact with other residents increased relative odds of medium to low level of satisfaction with housing conditions by a factor of 1.34 with 95% confidence interval (CI) of (1.04, 1.74) and increased relative odds of high to low level of satisfaction with housing conditions by a factor of 1.39, 95% CI: (1.10, 1.75). Given level of contact with other residents, comparing with apartment residents, subjects living in house had lower relative odds of high to low level of satisfaction with housing conditions, the odds ratio was 0.74, 95% CI: (0.57, 0.96), while the difference between house and apartment residents was not significant in the relative odds of medium to low level of satisfaction with housing conditions (1.07, 95% CI:(0.81, 1.42)). Compared with apartment residents, the tower block residents had a higher relative odds of medium to low level of satisfaction with housing condition by a factor of 1.50, 95% CI: (1.07, 2.10), and a higher relative odds of high to low level of satisfaction with housing condition by a factor of 1.90, 95% CI: (1.42, 2.55).

Hypotheses on interactions, $H_0 : \beta_{j4} = \beta_{j5} = 0; j = 2, 3$ vs. H_1 : Not all β_{j4}, β_{j5} are zero. Likelihood ratio test $T_S = 3605.48 - 3598.587 = 6.893, df = 4, p = 0.1417$. Wald test $T_S = 6.90, df = 4, p = 0.1413$. Null hypothesis of no interaction was not rejected. There was no interaction of house type and level of contact with other residents on the relative odds of medium or high level to the low level of satisfaction of housing conditions.

3. As the response has ordinal categories, fit proportional odds model to the data that include the same variables as used in the nominal logistic model obtained in (ii). What does the fitted model tell?
   Fit proportional odds model for the ordinal response of satisfaction (1: low, 2: Medium, 3: high).

$$\begin{aligned} \log\left(\frac{\gamma_j}{1-\gamma_j}\right) &= \alpha_j + \beta_1 C + \beta_2 TypeH + \beta_3 TypeT, j = 1, 2 \\ \gamma_1 &= \pi_1 \\ \gamma_2 &= \pi_1 + \pi_2 \\ \gamma_3 &= 1 \end{aligned}$$

	Statistic (df)	p - value
Score test for proportional odds assumption	4.85 (df = 3)	0.1833
Deviance	11.70 (df = 7)	0.1109
Pearson Chi-square	11.64 (df = 7)	0.1130
Wald test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0 (Contact)$	7.39 (df = 1)	0.0065
Wald test $H_0 : \beta_2 = \beta_3 = 0$ vs. $H_1 : \beta_2 \neq 0$ or $\beta_3 \neq 0 (HouseType)$	34.49 (df = 2)	< 0.0001

The proportional odds assumption seemed met and the model seemed to fit data well. The fitted model suggested that given type of housing, high level of contact with other residents increase the cumulative odds for higher level of satisfaction with housing conditions by a factor of 1.29 with 95% CI of (1.07, 1.54). For each level of contact, comparing to apartment residents, tower block residents had higher cumulative odds for higher level of satisfaction with housing conditions by a factor of 1.65 with 95% CI of (1.31, 2.07), while subjects living in house had lower cumulative odds than apartment residents for higher level of satisfaction with housing conditions, the odds ratio was 0.79 with 95% CI of (0.64, 0.97).

4. Calculate Pearson residuals from the proportional odds model for ordinal response to find where the largest discrepancies are between the observed frequencies and expected frequencies estimated from the model.

		Level of Satisfaction			Pearson residuals		
		Low	Medium	High	Level of Satisfaction		
Type	Contact	Observed (E)	Observed (E)	Observed (E)	Low	Medium	High
Tower	Low	65 (59.01)	54 (56.78)	100 (103.20)	.779	-.369	-.315
	High	34 (40.32)	47 (43.98)	100 (96.71)	-.995	.455	.335
Apt	Low	130 (119.95)	76 (85.89)	111 (111.16)	.918	-1.067	-.015
	High	141 (143.84)	116 (120.44)	191 (183.72)	-.237	-.405	.537
House	Low	67 (77.01)	48 (47.04)	62 (52.95)	-1.141	.140	1.244
	High	130 (126.90)	105 (91.89)	104 (120.21)	.275	1.368	-1.479

The largest discrepancies between the observed and expected frequencies estimated from the model were in the cells for those living in house, with high level of contact with other residents, and satisfaction levels of medium and high. The largest magnitude of Pearson residuals was less than 1.5, indicating that there were no outliers.

Appendix

```
## R code
library(dplyr)
library(MASS)
library(nnet)
```

```
dat <- c(65, 34, 54, 47, 100, 100,
         130, 141, 76, 116, 111, 191,
         67, 130, 48, 105, 62, 104)
dat_tb<- array(dat, c(2, 3, 3),
dimnames = list(Contact = c("Low", "High"),
            Response = c("Low.sat", "Median.sat", "High.sat"),
            Area = c("Tower", "Apartment", "House")))
ftable(dat_tb)

################################
## question 1
################################
## 1.1
hous <- margin.table(dat_tb, margin =c(2, 3))
prop.table(hous, margin = 2) %>% round(., 4)
chisq.test(hous)

## 1.2
hous <- margin.table(dat_tb, margin =c(1, 3))
prop.table(hous, margin = 2) %>% round(., 4)
chisq.test(hous)

## 1.3
cont <- margin.table(dat_tb, margin =c(1, 2))
prop.table(cont, margin = 1) %>% round(., 4)
chisq.test(cont)

################################
## question 2
################################
dat_full <- data.frame(low = as.vector(dat_tb[,1,]),
                       median = as.vector(dat_tb[,2,]),
                       high = as.vector(dat_tb[,3,]),
                       int = rep(c(1, 2), 3),
                       cate = rep(c(2, 1, 3), each = 2))

## 2.1
## model 1 without interaction
m1 <- multinom(cbind(dat_full$low, dat_full$median, dat_full$high)
~ factor(int) + factor(cate), data = dat_full)
summary(m1)
exp(coef(m1))
exp(confint(m1))
```

```
## model 2 with interaction
m2 <- multinom(cbind(dat_full$low, dat_full$median, dat_full$high)
~ factor(int) +  factor(cate) + factor(cate)*factor(int), data = dat_full)
summary(m2)
exp(coef(m2))
exp(confint(m2))

## LRT
TS1 <- deviance(m1) - deviance(m2)
p1 <- 1-pchisq(TS1, 4)

## model 3 with house type
m3 <- multinom(cbind(dat_full$low, dat_full$median,
dat_full$high) ~ factor(cate), data = dat_full)
TS2 <- deviance(m3) - deviance(m1)
p2 <- 1-pchisq(TS2, 2)

## model 4 with contact with others
m4 <- multinom(cbind(dat_full$low, dat_full$median,
dat_full$high) ~ factor(int), data = dat_full)
TS3 <- deviance(m4) - deviance(m1)
p3 <- 1-pchisq(TS3, 4)

## goodness of fit
pihat <- predict(m1, type = "probs")
m <- rowSums(dat_full[,1:3])
res.pearson <- (dat_full[,1:3]-pihat*m)/sqrt(pihat*m)
G.stat <- sum(res.pearson^2)
p4 <- 1-pchisq(G.stat, (6-4)*(3-1))

################################
## question 3
################################
freq <- c(dat_full$low, dat_full$median, dat_full$high)
res <- c(rep(c("L", "M", "H"), c(6, 6, 6)))
res <- factor(res, levels = c("L", "M", "H"), ordered = T)
dat_ord <- data.frame(res = res, int = rep(dat_full$int, 3),
cate = rep(dat_full$cate, 3), freq = freq)

m5 <- polr(res ~ factor(int) + factor(cate), dat_ord,
weights = freq, method = "logistic")
summary(m5)
exp(coef(m5))
```

```
exp(confint(m5))

m6 <- polr(res ~ factor(cate), dat_ord, weights = freq)
TS5 <- deviance(m6) - deviance(m5)
p5<- 1-pchisq(TS5, 1)

m7 <- polr(res ~ factor(int), dat_ord, weights = freq)
TS6 <- deviance(m7) - deviance(m5)
p6<- 1-pchisq(TS6, 2)

#Pearson Chi-Square residuals from proportional odds model
pihat <- predict(m5, type = "probs")
m <- rowSums(dat_full[,1:3])
res.pearson <- ((dat_full[,1:3]-pihat*m)/sqrt(pihat*m))
%>% round(.,2) %>% as.data.frame()
```