

HW4 yq2378

Yumeng Qi

2024-02-27

1 summarize th data

```
df1 = tibble(  
  satisfaction = rep(c("low", "medium", "high"), each = 6),  
  contact = rep(rep(c("low", "high"), each = 3), 3),  
  housing = rep(c("Tower Block", "Apartment", "House"), 6),  
  cnt = c(65, 130, 67, 34, 141, 130, 54, 76, 48, 47, 116, 105, 100, 111, 62, 100, 191, 104)) |>  
  mutate(satisfaction = factor(satisfaction, levels = c("low", "medium", "high"), ordered = TRUE),  
         contact = factor(contact, levels = c("high", "low")) )
```

```
# Create a table for satisfaction vs contact  
contact_table <- xtabs(cnt ~ satisfaction + contact, data = df1)
```

```
# Convert counts to percentages  
round(prop.table(contact_table, 2) * 100, 2)
```

```
##           contact  
## satisfaction high  low  
##      low      31.51 36.75  
##      medium 27.69 24.96  
##      high   40.81 38.29
```

```
round(prop.table(contact_table, 1) * 100, 2)
```

```
##           contact  
## satisfaction high  low  
##      low      53.79 46.21  
##      medium 60.09 39.91  
##      high   59.13 40.87
```

```
# Create a table for satisfaction vs housing  
housing_table <- xtabs(cnt ~ satisfaction + housing, data = df1)
```

```
# Convert counts to percentages  
round(prop.table(housing_table, 2) * 100, 2)
```

```
##           housing  
## satisfaction Apartment House Tower Block  
##      low           35.42 38.18         24.75  
##      medium        25.10 29.65         25.25  
##      high          39.48 32.17         50.00
```

1) contact with other residents and levels of satisfaction

High contact with other residents has a larger proportion of high satisfaction of residents (40.81%) compared to low contact with other residents (38.29%). However, medium satisfaction residents have the least proportion in both high and low contact with other.

2) type of housing and levels of satisfaction

Those living in the block has 50% of high satisfaction and nearly equal amount of low and medium satisfaction. While for those living in the house, the majority (38.18%) has low satisfaction.

2

```
df2 = tibble(
  contact = rep(c("low", "high"), each = 3),
  housing = rep(c("Tower Block", "Apartment", "House"), 2),
  low_satisfaction = c(65, 130, 67, 34, 141, 130),
  medium_satisfaction = c(54, 76, 48, 47, 116, 105),
  high_satisfaction = c(100, 111, 62, 100, 191, 104))

library(nnet)
mult <- multinom(cbind(low_satisfaction, medium_satisfaction, high_satisfaction) ~ contact + housing, data = df2)

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

summary_mult <- summary(mult)

summary_mult

## Call:
## multinom(formula = cbind(low_satisfaction, medium_satisfaction,
##   high_satisfaction) ~ contact + housing, data = df2)
##
## Coefficients:
##               (Intercept) contactlow housingHouse housingTower Block
## medium_satisfaction -0.2180364 -0.2959832  0.06967922      0.4067631
## high_satisfaction   0.2474047 -0.3282264  -0.30402275      0.6415948
##
## Std. Errors:
##               (Intercept) contactlow housingHouse housingTower Block
## medium_satisfaction  0.10930968  0.1301046   0.1437749      0.1713009
## high_satisfaction   0.09783068  0.1181870   0.1351693      0.1500774
##
## Residual Deviance: 3605.48
## AIC: 3621.48
```

For “medium satisfaction” relative to the baseline “low satisfaction”

Being low in contact with others has a negative coefficient (-0.2960), suggesting that lower contact is associated with a decrease in the log-odds of medium satisfaction. As for housing in House and Tower Block, both have positive coefficient (0.0697 and 0.4068 respectively), suggesting an increase in the log-odds of medium

satisfaction and clearly Tower Block has a stronger impact than House.

For “high satisfaction” relative to the baseline “low satisfaction”

Being low in contact with others has a negative coefficient (-0.3282), which means larger impact, but still indicates consistency in the direction with medium satisfaction. Housing in House has a negative coefficient (-0.3040), suggesting living in a house is associated with a decrease in the log-odds of high satisfaction. Housing in Tower Block has a positive coefficient, significantly increasing the log-odds of high satisfaction for tower block residents

```
coefficients <- summary_mult$coefficients
std_errors <- summary_mult$standard.errors

# Calculate 95% Confidence Intervals
conf_int_lower <- coefficients - 1.96 * std_errors
conf_int_upper <- coefficients + 1.96 * std_errors

CI = paste("(",round(conf_int_lower,2),", ", round(conf_int_upper,2), ")",sep = ",")

CI = rbind(CI[seq(1,8,by =2)],CI[seq(2,8,by =2)])
colnames(CI) = colnames(conf_int_lower)
rownames(CI) = rownames(conf_int_lower)
print("95% Confidence Interval")

## [1] "95% Confidence Interval"
CI

##               (Intercept)      contactlow      housingHouse
## medium_satisfaction "(-0.43, 0)"    "(-0.55, -0.04)" "(-0.21, 0.35)"
## high_satisfaction   "(0.06, 0.44)"  "(-0.56, -0.1)"  "(-0.57, -0.04)"
##               housingTower Block
## medium_satisfaction "(0.07, 0.74)"
## high_satisfaction   "(0.35, 0.94)"
```

From the CI, we find only housing being House and medium satisfaction has a CI contains 0. It indicates except for this combination, all other coefficients are significant under 0.05 significance value.

```
# goodness of fit
pihat1=predict(mult,type='probs')
m = rowSums(df2[,3:5])
res.pearson1=(df2[,3:5]-pihat1*m)/sqrt(pihat1*m) # pearson residuals
G.stat1=sum(res.pearson1^2) # Generalized Pearson Chisq Stat
G.stat1 # Generalized Pearson chisq statistic

## [1] 6.932341

D.stat1=sum(2*df2[,3:5]*log(df2[,3:5]/(m*pihat1)))
D.stat1 # Deviance statistics
```

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

As for the goodness of fit, the Generalized Pearson chisq statistic is 6.9323412 and the Deviance statistics is 6.8930277.

```
mult2 <- multinom(cbind(low_satisfaction, medium_satisfaction, high_satisfaction)~contact*housing, data)

## # weights: 21 (12 variable)
## initial value 1846.767257
## iter 10 value 1800.128659
```

```
## final value 1799.293647
## converged
```

```
#summary(mult2)
anova(mult, mult2, test="Chisq")
```

```
##           Model Resid. df Resid. Dev   Test      Df LR stat.   Pr(Chi)
## 1 contact + housing           4    3605.480      NA         NA         NA
## 2 contact * housing           0    3598.587 1 vs 2      4  6.893028 0.1416504
```

Given interaction leads to a nested model, we test if the null model stands, which suggests that there is no interaction term. From the result we know the p-value is 0.1417. Under significance value of 0.05, we fail to reject the null hypothesis, believing there is no interaction of contact level by house type.

3

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##      select
```

```
# fit proportional odds model
polr= polr(satisfaction~contact+housing,data=df1,weights=cnt)
summary(polr)
```

```
##
## Re-fitting to get Hessian
## Call:
## polr(formula = satisfaction ~ contact + housing, data = df1,
##      weights = cnt)
##
## Coefficients:
##              Value Std. Error t value
## contactlow      -0.2524    0.09306  -2.713
## housingHouse     -0.2353    0.10521  -2.236
## housingTower Block  0.5010    0.11675   4.291
##
## Intercepts:
##              Value Std. Error t value
## low|medium    -0.7488    0.0818  -9.1570
## medium|high   0.3637    0.0801   4.5393
##
## Residual Deviance: 3610.286
## AIC: 3620.286
```

- contact low: This coefficient (0.2524) represents the effect of the “contact” predictor on the log odds of being in a higher satisfaction category, holding other variables constant. The positive sign indicates that being contact low is associated with lower odds of being in a higher satisfaction category compared to being contact high. This trend is consistent with nominal regression. The t-value of -2.713 and the corresponding standard error suggest that this effect is statistically significant.
- housing House: The coefficient (0.2353) for “housing House” indicates that living in a house, compared

to the baseline housing category, Apartment and among those with same level of contact, is associated with lower odds of being in a higher satisfaction category. The negative sign denotes a decrease in the odds of higher satisfaction for residents in houses, and the t-value of -2.236 suggests this effect is statistically significant. This is obviously not consistent with the nominal regression, as whose result shows non-monotone trend.

- housing Tower Block: The coefficient (-0.5010) for “housing Tower Block” suggests that living in a tower block, relative to the baseline, and among those with same level of contact, significantly increases the odds of being in a higher satisfaction category. This is consistent with the nominal regression. The positive coefficient and the t-value of 4.291 strongly indicate this is a significant predictor of higher satisfaction levels.

4 the largest discrepancies are between the observed frequencies and the expected frequencies.

```
# residuals
pihat=predict(polr,df1,type='p')
m=rowSums(cbind(df2$low_satisfaction,df2$medium_satisfaction,df2$high_satisfaction))
res.pearson=(df2[,3:5]-pihat*m)/sqrt(pihat*m)
G=sum(res.pearson^2)

which(abs(res.pearson) == max(abs(res.pearson)), arr.ind = TRUE)

##      row col
## [1,]    4   3
df2[4,]

## # A tibble: 1 x 5
##   contact housing    low_satisfaction medium_satisfaction high_satisfaction
##   <chr>   <chr>          <dbl>             <dbl>             <dbl>
## 1 high   Tower Block      34                47                100
```

The Pearson chi-square statistic is 181.6268587 and the high satisfaction with high contacts and housing in Tower Block has the largest discrepancies.