P8131_hw4_xy2395

Jack Yan

2/28/2019

Data Entry

```
data_wide = tibble(
  resp_low = c(65, 34, 130, 141, 67, 130),
  resp_medium = c(54, 47, 76, 116, 48, 105),
  resp_high = c(100, 100, 111, 191, 62, 104),
  contact = rep(c('low', 'high'), 3),
  type = c('tower_block', 'tower_block', 'apartment', 'apartment', 'house', 'house')
)

data_long =
  data_wide %>%
  gather(key = 'satisfaction', value = 'count', resp_low:resp_high) %>%
  mutate(satisfaction = str_sub(satisfaction, 6),
         satisfaction = fct_relevel(satisfaction, 'low', 'medium', 'high'),
         type = fct_relevel(type, 'tower_block','apartment', 'house'),
         contact = fct_relevel(contact, 'low', 'high')
  )

data_wide =
  data_long %>%
  spread(satisfaction, count) %>%
  mutate(
    contact = as_factor(contact),
    type = as_factor(type)
  ) %>%
  select(low, medium, high, everything())
```

i) Tables

```
tbl1 =
  data_long %>%
  group_by(contact, satisfaction) %>%
  summarize(sum = sum(count)) %>%
  spread(contact, sum) %>%
  mutate(high = 100 * high / sum(high),
         low = 100 * low / sum(low))

tbl1 %>% knitr::kable(digits = 2,
                      col.names = c('Satisfaction', 'Low Contact (%)', 'High Contact (%)'))
```

Satisfaction	Low Contact (%)	High Contact (%)
low	36.75	31.51
medium	24.96	27.69

Satisfaction	Low Contact (%)	High Contact (%)
high	38.29	40.81

The table above shows association between satisfaction and contact with others. Residents with higher degree of contact with others tend to be more satisfied with their housing conditions. The degree of contact is positively associated with housing satisfaction.

```
tbl2 =
  data_long %>%
  group_by(type, satisfaction) %>%
  summarize(sum = sum(count)) %>%
  spread(type, sum) %>%
  mutate(tower_block = tower_block / sum(tower_block),
         apartment = apartment / sum(apartment),
         house = house / sum(house)
  )

tbl2 %>%
  knitr::kable(digits = 2,
               col.names = c('Satisfaction', 'Tower Block (%)', 'Apartment (%)', 'House (%)'))
```

Satisfaction	Tower Block (%)	Apartment (%)	House (%)
low	0.25	0.35	0.38
medium	0.25	0.25	0.30
high	0.50	0.39	0.32

The table above shows association between satisfaction and type of housing. Tower block Residents have the highest proportion of high satisfaction (50%) and the lowest proportion of low satisfaction (25%). Most residents living in apartments have high level of satisfaction, whereas most residents living in houses have low level of satisfaction.

ii) Multinomial Logistic Regression

```
fit_nomial = multinom(satisfaction ~ contact + type, weights = count, data = data_long)
```

```
## # weights:  15 (8 variable)
## initial  value 1846.767257
## iter  10 value 1803.278543
## final  value 1802.740161
## converged
```

Pattern in the associations

```
# Coefficients
summary(fit_nomial)$coefficients
```

```
##        (Intercept) contacthigh typeapartment  typehouse
## medium  -0.1072644   0.2959803    -0.4067537 -0.3370771
## high     0.5607737   0.3282263    -0.6415967 -0.9456177
```

The log odds ratio of medium vs low satisfaction between high and low level of contact is 0.296. The log odds ratio of high vs low satisfaction between high and low level of contact is 0.323. This suggests that residents with higher levels of contact have higher level of satisfaction.

The log odds ratio of medium vs low satisfaction between apartment and tower block residents is -0.41. The log odds ratio of high vs low satisfaction between apartment and tower block residents is -0.64. This suggests that people living in tower blocks tends to have the highest level of satisfaction.

Odds ratios with 95% confidence intervals

```
fit_nomial %>%
  broom::tidy() %>%
  mutate(CIL = estimate * exp(std.error * qnorm(0.025)),
         CIR = estimate / exp(std.error * qnorm(0.025))) %>%
  select(-std.error, -statistic, -p.value) %>%
  filter(term != '(Intercept)') %>%
  knitr::kable(digits = 2)
```

y.level	term	estimate	CIL	CIR
medium	contacthigh	1.34	1.04	1.73
medium	typeapartment	0.67	0.48	0.93
medium	typehouse	0.71	0.50	1.02
high	contacthigh	1.39	1.10	1.75
high	typeapartment	0.53	0.39	0.71
high	typehouse	0.39	0.28	0.54

The odds ratios along with their 95% confidence intervals are shown in the table above.

Goodness of fit

```
# Goodness of fit
pi_hat = predict(fit_nomial, data_wide, type = 'probs')
# data_long
res = data_wide[,1:3]
m = rowSums(res)
G_stat = sum((res - pi_hat * m)^2 / (pi_hat * m))
G_stat
```

```
## [1] 6.932334
```

```
p_value = 1 - pchisq(G_stat, df = (6-4)*(3-1))
p_value
```

```
## [1] 0.1395076
```

The p-value is $0.1395076 > 0.05$, so the fit is good.

Testing interaction of contact level by house type

```
fit_nomial_interact = multinom(satisfaction ~ contact + type + contact*type, weights = count, data = da
```

```
## # weights:  21 (12 variable)
## initial  value 1846.767257
## iter  10 value 1800.614138
```

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

```
summary(fit_nomial_interact)
```

```
## Call:
## multinom(formula = satisfaction ~ contact + type + contact *
##     type, data = data_long, weights = count)
##
## Coefficients:
##        (Intercept) contacthigh typeapartment  typehouse
## medium  -0.1854029   0.5091910    -0.3513970 -0.1480892
## high     0.4307831   0.6480263    -0.5887874 -0.5083426
##        contacthigh:typeapartment contacthigh:typehouse
## medium                -0.1675605            -0.3892720
## high                  -0.1865089            -0.7936097
##
## Std. Errors:
##        (Intercept) contacthigh typeapartment typehouse
## medium   0.1841280   0.2908461     0.2339943 0.2639353
## high     0.1593255   0.2545513     0.2051486 0.2375690
##        contacthigh:typeapartment contacthigh:typehouse
## medium                 0.3480726             0.3708993
## high                   0.3063095             0.3363904
##
## Residual Deviance: 3598.587
## AIC: 3622.587
```

```
D.stat_small = sum(2 * res * log(res/(m*pi_hat)) )
D.stat_small
```

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

```
pi_hat2 = predict(fit_nomial_interact, data_wide, type = 'probs')
D.stat_large = sum(2 * res * log(res/(m*pi_hat2)) )
D.stat_large
```

```
## [1] 2.981992e-10
```

```
# Deviance analysis
pval = 1 - pchisq(D.stat_small - D.stat_large, df = 4)
pval
```

```
## [1] 0.1416504
```

We used deviance analysis to test the interaction between level of contact and housing type. The p-value is larger than 0.05, so we fail to reject the null and conclude that the smaller model (without the interaction) is better. In other words, there is no interaction between contact level and house type.

iii) Proportional odds model

```
fit_polr = MASS::polr(satisfaction ~ contact + type, data = data_long, weights = count)

summary(fit_polr)
```

```
##
```

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

```
## Call:
## MASS::polr(formula = satisfaction ~ contact + type, data = data_long,
##     weights = count)
##
## Coefficients:
##                 Value Std. Error t value
## contacthigh    0.2524    0.09306   2.713
## typeapartment -0.5009    0.11675  -4.291
## typehouse     -0.7362    0.12610  -5.838
##
## Intercepts:
##             Value   Std. Error t value
## low|medium  -0.9973  0.1075    -9.2794
## medium|high  0.1152  0.1047     1.1004
##
## Residual Deviance: 3610.286
## AIC: 3620.286
```

The log odds ratio between high and low levels of contact is 0.2524, so people with higher level of contact with others have greater satisfaction.

The log odds ratio between the housing types apartment and tower block is -0.5009. The log odds ratio between house and tower block is -0.7362. This tells us that house residents have the lowest level of satisfaction and tower block residents have the highest level of satisfaction.

iv) Pearson residuals

```
pihat_polr = predict(fit_polr, data_wide[,4:5], type = 'prob')

resid_matrix = cbind(data_wide[,4:5], (res - pihat_polr*m)/sqrt(pihat_polr*m))
resid_matrix
```

```
##   contact        type        low     medium        high
## 1     low tower_block  0.7793957 -0.3697193 -0.31511792
## 2     low   apartment  0.9177560 -1.0671823 -0.01527344
## 3     low       house -1.1407855  0.1397563  1.24407710
## 4    high tower_block -0.9946852  0.4549302  0.33544295
## 5    high   apartment -0.2369309 -0.4052334  0.53777345
## 6    high       house  0.2743817  1.3677881 -1.47782697
```

```
max(abs(resid_matrix[,3:5]))
```

```
## [1] 1.477827
```

The largest Pearson residual is -1.478, corresponding to the response level 'high satisfaction' and covariate levels 'high contact' and 'house'.