BM_HW4

Siyan Chen_sc4456

3/4/2019

data input

```
df = data.frame(ls_y = c(65, 130, 67, 34, 141, 130), ms_y = c(54, 76, 48, 47, 116, 105), hs_y = c(100, 
df$contact = factor(df$contact, level = c("low", "high"))
df$type = factor(df$type, level = c("tower_block", "apartment", "house"))
```

#1 Pair-wise Table

contact	ls	ms	hs	ls_p	ms_p	hs_p
low	262	178	273	36.75%	24.96%	38.29%
high	305	268	395	31.51%	27.69%	40.81%

type	ls	ms	hs	ls_p	ms_p	hs_p
tower_block	99	101	200	24.8%	25.25%	50.0%
apartment	271	192	302	35.4%	25.10%	39.5%
house	197	153	166	38.2%	29.65%	32.2%

ls_y	ms_y	hs_y	contact	type	ls_p	ms_p	hs_p
65	54	100	low	tower_block	29.7%	24.66%	45.7%
130	76	111	low	apartment	41.0%	23.97%	35.0%
67	48	62	low	house	37.9%	27.12%	35.0%
34	47	100	high	tower_block	18.8%	25.97%	55.2%
141	116	191	high	apartment	31.5%	25.89%	42.6%
130	105	104	high	house	38.3%	30.97%	30.7%

According to the tables, when residents have high degree contact, the proportion of response in medium or high satisfaction categories tend to increase. When residents lives in tower block, the proportion of response in high satisfaction categorys tend to increase. When residents lives in tower block and have high degree contact, the proportion of response in high satisfaction category is greatest (55.2%).

2 Nomial logisitic regression

```
multi_model = multinom(cbind(ls_y, ms_y, hs_y) ~ contact + type, data = df)
```

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

```r
summary(multi_model)
```

```
## Call:
## multinom(formula = cbind(ls_y, ms_y, hs_y) ~ contact + type,
##     data = df)
##
## Coefficients:
##      (Intercept) contacthigh typeapartment  typehouse
## ms_y  -0.1072644   0.2959803    -0.4067537 -0.3370771
## hs_y   0.5607737   0.3282263    -0.6415967 -0.9456177
##
## Std. Errors:
##      (Intercept) contacthigh typeapartment typehouse
## ms_y   0.1524077   0.1301046     0.1713011 0.1803577
## hs_y   0.1329301   0.1181870     0.1500773 0.1644850
##
## Residual Deviance: 3605.48
## AIC: 3621.48
```

Interpretation: $log\frac{\pi_{medium}}{\pi_{low}} = \beta_{10}+\beta_{11}*I\{contact = high\}+\beta_{12}*I\{type = apartment\}+\beta_{13}*I\{type = house\}$

β_{10}: Log odds for response of Medium satisfaction versus Low satisfcation is -0.107 if residents live in tower block and have low contacts with others.

β_{11}: Log odds for response of Medium satisfaction versus Low satisfcation increases 0.296 if residents have high contacts while house type does not changes.

β_{12} ; Log odds for response of Medium satisfaction versus Low satisfcation decreases 0.407 if residents live in apartment while contact frequency does not change.

β_{13}Log odds for response of Medium satisfaction versus Low satisfcation decreases 0.337 if residents live in house while contact frequency does not change.

2. $log\frac{\pi_{high}}{\pi_{low}} = \beta_{20} + \beta_{21} * I\{contact = high\} + \beta_{22} * I\{type = apartment\} + \beta_{23} * I\{type = house\}$

β_{20}:Log odds for response of High satisfaction versus Medium satisfcation is 0.561 if residents live in tower block and have low degree of contacts with others.

β_{21}:Log odds for response of High satisfaction versus Medium satisfcation increases 0.328 if residents have high degree of contacts while house type does not changes.

β_{22}:Log odds for response of High satisfaction versus Medium satisfcation decreases 0.642 if residents live in apartment while degree of contact does not change.

β_{23}:Log odds for response of High satisfaction versus Medium satisfcation decreases 0.946 if residents live in house while degree of contact does not change

95% CI

```r
beta = summary(multi_model)$coefficients
se1=sqrt(vcov(multi_model)[1,1])
se2 = sqrt(vcov(multi_model)[2,2])
se3=sqrt(vcov(multi_model)[3,3])
se4=sqrt(vcov(multi_model)[4,4])
se5=sqrt(vcov(multi_model)[5,5])
se6=sqrt(vcov(multi_model)[6,6])
```

```r
se7=sqrt(vcov(multi_model)[7,7])
se8=sqrt(vcov(multi_model)[8,8])
exp(beta[1]+c(qnorm(0.025),0,-qnorm(0.025))*se1)
exp(beta[2]+c(qnorm(0.025),0,-qnorm(0.025))*se5)
exp(beta[3]+c(qnorm(0.025),0,-qnorm(0.025))*se2)
exp(beta[4]+c(qnorm(0.025),0,-qnorm(0.025))*se6)
exp(beta[5]+c(qnorm(0.025),0,-qnorm(0.025))*se3)
exp(beta[6]+c(qnorm(0.025),0,-qnorm(0.025))*se7)
exp(beta[7]+c(qnorm(0.025),0,-qnorm(0.025))*se4)
exp(beta[8]+c(qnorm(0.025),0,-qnorm(0.025))*se8)
```

model1: CI β_{10}: [0.6663249 0.8982882 1.2110033]
β_{11}: [1.041831 1.344444 1.734954]
β_{12}: [0.4759238 0.6658082 0.9314528]
β_{13}: [0.5012894 0.7138538 1.0165531]

We are 95% confident that odds for response of Medium satisfaction versus Low satisfcation is between 0.666 and 1.211 if residents live in tower block and have low contacts with others.

We are 95% confident that odds ratio for response of Medium satisfaction versus Low satisfcation for high contact vs low contact is between 1.042 and 1.735.

We are 95% confident that odds ratio for response of Medium satisfaction versus Low satisfcation for apartment vs tower_block is between 0.476 and 0.931.

We are 95% confident that odds ratio for response of Medium satisfaction versus Low satisfcation for house vs tower_block is between 0.501 and 1.017.

Simialr explaination for model 2

model2: CI β_{20}: [1.350177 1.752027 2.273480]
β_{21}: [1.101402 1.388503 1.750442]
β_{22}: [0.3922944 0.5264512 0.7064869]
β_{23}: [0.2813932 0.3884396 0.5362080]

We are 95% confident that odds for response of Medium satisfaction versus High satisfcation is between 1.35 and 2.27 if residents live in tower block and have low contacts with others.

goodness of fit

```r
pihat = predict(multi_model, type = "probs")
m = rowSums(df[,1:3])
res_pearson  = (df[,1:3] - pihat*m)/sqrt(pihat*m) ## pearson residuals
G.stats = sum(res_pearson^2)
G.stats
```

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

```r
pval = 1-pchisq(G.stats, df = (3-1)*(6-3))
pval # could not reject the null hypothesis, fit well
```

```
## [1] 0.3271503
```

The pearson deiviance of model is 6.932334, which follows chi_square distribution with df = (3-1)*(6-3). According p value, this model fits data well.

3 Proportion Model

```
freq = c(df$ls_y, df$ms_y, df$hs_y)
res = c(rep(c("ls", "ms", "hs"), c(6, 6, 6)))
res = factor(res, level = c("ls", "ms", "hs"), ordered = T)
df2 = data.frame(res, contact = c("low", "low", "low", "high","high","high"), type = c(rep(c("tower_blo
df2$type = factor(df2$type, level = c("tower_block", "apartment", "house" ))
df2$contact = factor(df2$contact, level = c("low","high"))
```

```
# fit proportional odds model
polr_model = polr(res ~ contact + type, data = df2, weights = freq)
summary(polr_model)
```

```
## Call:
## polr(formula = res ~ contact + type, data = df2, weights = freq)
##
## 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
## ls|ms -0.9973  0.1075    -9.2794
## ms|hs  0.1152  0.1047     1.1004
##
## Residual Deviance: 3610.286
## AIC: 3620.286
```

$log\frac{P(Y\leq j)}{P(Y>j)} = aj + 0.2524 * I\{contact = high\} - 0.5009 * I\{type = apartment\} - 0.7362 * I\{type = house\}$

ls|ms -0.9973 is a1 ms|hs 0.1152 is a2

aj: the log odds of falling into or below category j when contact = low and type = tower block

a2: The estimated log odds of response falls in medium satisfaction or lower is 0.1152 when contact = low and type = tower block

β_k: incrrease in log-odds of falling into or below category with one unit increase in x_k holding all other x vaiables constant.

$\beta_1 = 0.2524$: Increase in log-odds of falling into or below category when contact changes to high while other x variables constant.

#4

```
pihat1 = predict(polr_model,df,type='p')
res.pearson=(df[,1:3] - pihat1*m)/sqrt(pihat1*m)
G=sum(res.pearson^2)

p_polr = predict(polr_model, df, type = 'p')
res.pearson_2 = (df[,1:3] - p_polr*m)/sqrt(p_polr*m)

abs(res.pearson_2)[6,3]
```

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

According to the pearson residuals for matrix, the greatest discrepancies between the observed frequencies and expected frequencies estimated from the model occurs in category of high frequency of contact and house type.