# Lec_23_2

## Exercise 12

```
## Data and model
library(dplyr)
library(broom)
```

```
## Warning: package 'broom' was built under R version 3.5.2
```

```
library(car)
library(multcomp)
```

```
## Warning: package 'multcomp' was built under R version 3.5.3
```

```
## Warning: package 'mvtnorm' was built under R version 3.5.2
```

```
## Warning: package 'TH.data' was built under R version 3.5.3
```

```
poli <- read.csv("PolIdeolData.csv")
poli$ideol <- factor(poli$ideol, levels=c("VL", "SL", "M", "SC", "VC"))
lin.score1 <- c(rep(x = c(1,2,3,4,5), times = 4))
poli %>% mutate(lin.score1 = lin.score1) -> poli
s.fit <- glm(count ~ gender*party + gender*ideol + party*lin.score1, family = poisson(link = "log"), data = poli)
```

a.

```
## Create new sets of scores
lin.score2 <- c(rep(c(0,2,3,4,6), 4))
lin.score3 <- c(rep(c(2,1,0,1,2), 4))
poli %>% mutate(lin.score2 = lin.score2, lin.score3 = lin.score3) -> poli

## Fit the model with new scores
s.fit.2 <- glm(count ~ gender*party + gender*ideol + party*lin.score2, family = poisson(link = "log"), data = poli)
s.fit.3 <- glm(count ~ gender*party + gender*ideol + party*lin.score3, family = poisson(link = "log"), data = poli)

## Compare p-values of PI interactions from the 3 models
anv.s <- tidy(Anova(s.fit))
anv.s2 <- tidy(Anova(s.fit.2))
anv.s3 <- tidy(Anova(s.fit.3))
anv.s$p.value[7] -> score1
anv.s2$p.value[7] -> score2
anv.s3$p.value[7] -> score3
data.frame(p.value = rbind(score1, score2, score3))
```

```
##             p.value
## score1 9.824709e-14
## score2 9.986882e-13
## score3 6.528893e-01
```

Hypothesis test:

H0: Interaction between party and ideology score has no effect; Ha: Interaction between party and ideology score has effect

use alpha = 0.05

P-value: 9.824709e-14 for score1, 9.986882e-13 for score2, 6.528893e-01 for score3.

Reject H0 when score is {0,2,3,4,6} and {1,2,3,4,5}, do not reject when score is {2,1,0,1,2}.

One can see H0 is rejected in the first 2 models, and p-value are close, so we can say {0,2,3,4,6} fits as well as {1,2,3,4,5}.

b.

```
## Fit models with GI score interactions
ss.fit <- glm(count ~ gender*party + ideol + gender*lin.score1 + party*lin.score1, family = poisson(link = "log"), data = poli)
ss.fit2 <- glm(count ~ gender*party + ideol + gender*lin.score2 + party*lin.score1, family = poisson(link = "log"), data = poli)
ss.fit3 <- glm(count ~ gender*party + ideol + gender*lin.score3 + party*lin.score1, family = poisson(link = "log"), data = poli)

## Deviances
GI.score1 <- ss.fit$deviance
GI.score2 <- ss.fit2$deviance
GI.score3 <- ss.fit3$deviance
original.model <- s.fit$deviance
data.frame(Deviance = rbind(original.model, GI.score1, GI.score2, GI.score3))
```

```
##                  Deviance
## original.model   8.398557
## GI.score1       16.842309
## GI.score2       17.095966
## GI.score3       11.628227
```

i. We can see that when parameters are less residual deviance is more for each model. That is, these new models with fewer parameters have less abilities to explain the association.

ii. Test siginificance of GI term

```
## Anvoa
i <- tidy(Anova(ss.fit))
ii <- tidy(Anova(ss.fit2))
iii <- tidy(Anova(ss.fit3))
data.frame(p.value = rbind(i$p.value[6], ii$p.value[6], iii$p.value[6]))
```

```
##      p.value
## 1 0.34491563
## 2 0.04274314
## 3 0.02594774
```

Hypothesis test:

H0: GI is has parameter 0; Ha: GI's parameter is not 0

alpha = 0.05

P.value: Refer to the table above.

We reject H0 only in the 2nd and 3rd model.

Therefore, GI association is signifiacant only when score is {2,1,0,1,2} and {0,2,3,4,6}.

c.

```
## Final model is model#3 with 11.63 deviance
fin.fit <- ss.fit3

## Remove missing value
fin.fit$coefficients <- na.omit(fin.fit$coefficients)

## See the ordering of parameters
coef(fin.fit)
```

```
##         (Intercept)            genderM             partyR 
##          3.79744833        -0.66793411        -1.53027724 
##             ideolSL             ideolM            ideolSC 
##          0.09449776         0.91538973        -0.33410712 
##             ideolVC     genderM:partyR genderM:lin.score3 
##         -0.49734923         0.31292785         0.20841353 
##  partyR:lin.score1 
##          0.43342092 
## attr(,"na.action")
## lin.score3 lin.score1 
##          8          9 
## attr(,"class")
## [1] "omit"
```

```
## Coefficient matrix
K <- matrix(c(rep(0, 8), 1, 1), nrow = 1, byrow = TRUE)

## Estimate OR and CI
wald.gi <- glht(fin.fit, K)
wald.ci.gi <- round(exp(confint(wald.gi, calpha = qnorm(0.975))$confint), 2)
row.names(wald.ci.gi) <- c("OR")
wald.ci.gi
```

```
##    Estimate  lwr  upr
## OR      1.9 1.55 2.33
## attr(,"conf.level")
## [1] 0.95
## attr(,"calpha")
## [1] 1.959964
```

i.

Odds ratios is 1.9 with confidence interval [1.55, 2.33]. This odds ratio indicates by how much the mean ratio between genders changes multiplicatively when ideol progressively becomes more conservative.

ii.

By comparing the nominal and ordinal estimates, the estimated ORs and CIs are not similar. As for CIs, the 2 types of CIs do not overlap. They do not have similar interpretations, ordinal variable requires a c-unit change in score have the same effect regardless of the initial score, while the nominal variable does not have such a restriction. And because in the model with ordinal variable we omitted many parameters from full model, the variance becomes smaller(bias-variance trade-off), theerefore CI gets narrower.

d.

We can see that gender and extremeness of ideology are siginificantly correlated. In particular, when ideology becomes 1 unit more extreme the male vs female mean ratio increses by 1.23 times, and the confidence interval indicates this multiplicative change always exists(CI doesn’t cover 1).

It is important to choose the proper set of scores, we can make reasonable guess of the relationship that scores measure(e.g. extremeness, conservativeness…etc) before fit the model.

What’s important when choosing scores is the progressive changes in the scores, because mean ratios and odds ratios are affected only by the difference between 2 scores, raw values doesn’t matter that much.

By adopting scores in the interaction terms we give up many parameters. Comapre to nominal estiamtes, in this way the inference seems to be more accurate(narrow CI), but the model has less ability to explain the deviance(large residual deviance).