

Applied Categorical Data Analysis: Homework 6

Kerem Tuncer

04/10/2021

Question 1

```
rm(list = ls())
```

Part A

We should keep in mind that a positive slope in the cumulative logit model (with the proportional odds property) indicates a negative association between x and the ordinal variable Y , whereas a negative slope indicates a positive association between x and the ordinal variable Y .

For x_1 , we have $\beta < 0$. Hence, the response level increases with an increase in x_1 . This means that, job satisfaction tends to increase at higher x_1 , which is earnings compared to others with similar positions.

For x_2 , we have $\beta > 0$. Hence, the response level increases with a decrease in x_2 . This means that, job satisfaction tends to increase at lower x_2 , which is freedom to make decisions.

For x_3 , we have $\beta > 0$. Hence, the response level increases with a decrease in x_3 . This means that, job satisfaction tends to increase at lower x_3 , which is work environment productivity.

These interpretations all make sense logically speaking.

Part B

For x_1 , we have $\beta < 0$. Hence, there is a positive association, meaning that Y will be the highest (which is 4) when x_1 is the highest which is $x_1 = 4$ (much more earnings compared to others with similar positions).

For x_2 , we have $\beta > 0$. Hence, there is a negative association, meaning that Y will be the highest (which is 4) when x_2 is the lowest which is $x_2 = 1$ (very true about being free to make decisions).

For x_3 , we have $\beta > 0$. Hence, there is a negative association, meaning that Y will be the highest (which is 4) when x_3 is the lowest which is $x_3 = 1$ (strongly agree about work environment allowing productivity).

Part C

Given that $\text{logit}[\hat{P} \geq j]$ is the reciprocal of $\text{logit}[\hat{P} \leq j]$, we can say that the signs of the slopes for x_1 , x_2 , and x_3 will change. The slope of x_1 will become positive, and the slopes of x_2 and x_3 will become negative.

Question 2

Part A

```
library(VGAM)

## Loading required package: stats4

## Loading required package: splines

Not <- c(6,6,6);
Pretty <- c(43,113,57);
Very <- c(75,178,117);
Income <- c("Below", "Average", "Above")
data.frame(Income, Not, Pretty, Very)

##      Income Not Pretty Very
## 1   Below    6    43    75
## 2 Average    6   113   178
## 3   Above    6    57   117

scores<-c(1,2,3)
logitmodel <- vglm(cbind(Not,Pretty,Very) ~ scores,
family=cumulative(parallel=TRUE))
summary(logitmodel)

##
## Call:
## vglm(formula = cbind(Not, Pretty, Very) ~ scores, family =
cumulative(parallel = TRUE))
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  -3.2466      0.3404  -9.537  <2e-16 ***
## (Intercept):2  -0.2378      0.2592  -0.917    0.359
## scores          -0.1117      0.1179  -0.948    0.343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])
##
## Residual deviance: 3.2472 on 3 degrees of freedom
##
## Log-likelihood: -15.4146 on 3 degrees of freedom
##
## Number of Fisher scoring iterations: 4
```

```
##
## No Hauck-Donner effect found in any of the estimates
##
##
## Exponentiated coefficients:
##   scores
## 0.8942746
```

The prediction equations are the following:

For $j = 1$,

$$\text{logit}[P(\hat{Y} \leq 1)] = -3.2466 - 0.1117 * \text{score}$$

For $j = 2$,

$$\text{logit}[P(\hat{Y} \leq 2)] = -0.2378 - 0.1117 * \text{score}$$

Part B

```
exp(-coef(logitmodel)[3])

##   scores
## 1.118225
```

For any fixed j , a unit increase in income score (from a lower score group to a higher score group) will increase the odds of being happier by a multiplicative effect of $\exp(0.1117423)=1.118225$.

Part C

Null hypothesis: $\beta = 0$, meaning that there is no effect of income on marital happiness. In other words, the two things are independent.

Null hypothesis: $\beta \neq 0$, meaning that there is an effect of income on marital happiness. In other words, the two things are not independent.

```
logit.null <- update(logitmodel, ~ 1)
lrtest(logitmodel, logit.null)

## Likelihood ratio test
##
## Model 1: cbind(Not, Pretty, Very) ~ scores
## Model 2: cbind(Not, Pretty, Very) ~ 1
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    3 -15.415
## 2    4 -15.858  1 0.8876    0.3461
```

Our test statistic is equal to 0.8876 and p-value of 0.3461. Given that the p-value is much larger than 0.05, we do not have enough evidence to conclude that income has an effect on marital happiness. This is in line with our findings from the previous homework.

Part D

```
summary(logitmodel)

##
## Call:
## vglm(formula = cbind(Not, Pretty, Very) ~ scores, family =
## cumulative(parallel = TRUE))
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  -3.2466    0.3404  -9.537  <2e-16 ***
## (Intercept):2  -0.2378    0.2592  -0.917    0.359
## scores          -0.1117    0.1179  -0.948    0.343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])
##
## Residual deviance: 3.2472 on 3 degrees of freedom
##
## Log-likelihood: -15.4146 on 3 degrees of freedom
##
## Number of Fisher scoring iterations: 4
##
## No Hauck-Donner effect found in any of the estimates
##
##
## Exponentiated coefficients:
##      scores
## 0.8942746
```

The residual deviance is 3.2472 and the degrees of freedom is 3. With this information, we can compute a deviance goodness-of-fit test.

```
1-pchisq(3.2472, 3)
```

```
## [1] 0.3550593
```

H0: model fits adequately

Ha: model does not fit adequately

As a result, we get the p-value of 0.3550593. Given that the p-value is above 0.05, we can say that the model fits adequately.

Part E

```
predict(logitmodel, data.frame(scores = 2), type="response")

##           Not      Pretty      Very
## 1 0.03017419 0.3565176 0.6133082
```

The probability that a person with average family income (score = 2) reports a very happy marriage is 0.6133082.

Question 3

Part A

```
Heaven_Hell <- matrix(c(833,125,2,160),ncol=2,byrow=TRUE)
colnames(Heaven_Hell) <- c("Hell (Yes)","Hell (No)")
rownames(Heaven_Hell) <- c("Heaven (Yes)","Heaven (No)")
Heaven_Hell <- as.table(Heaven_Hell)
Heaven_Hell
```

| ## | Hell (Yes) | Hell (No) |
|-----------------|------------|-----------|
| ## Heaven (Yes) | 833 | 125 |
| ## Heaven (No) | 2 | 160 |

$$H_0: P(Y_1 = 1) = P(Y_2 = 1)$$

In words, that would be that the population proportions answering yes were the same for heaven and hell.

$$H_a: P(Y_1 = 1) \neq P(Y_2 = 1)$$

In words, that would be that the population proportions answering yes were not the same for heaven and hell.

```
mcnemar.test(Heaven_Hell, correct = F)
```

```
##
## McNemar's Chi-squared test
##
## data: Heaven_Hell
## McNemar's chi-squared = 119.13, df = 1, p-value < 2.2e-16
```

Our chi-squared statistic is 119.3 and $df = 1$, so our p-value is $< 2.2e-16$. Given that this value is much below 0.05, we can reject the null hypothesis. We have enough to conclude that the population proportions answering yes were not the same for heaven and hell.

Part B

```
SE <- 1/1120 * sqrt( 125 + 2 - (125-2)^2 / 1120)
SE
```

```
## [1] 0.00951184
```

```
CI <- ((125 - 2) / 1120) + c(-1,1) * 1.64 * SE
CI
```

```
## [1] 0.09422201 0.12542085
```

The standard error was 0.00951184. Therefore, our 90% confidence interval is [0.09422201, 0.12542085]. We are 90% confident that the difference between the

population proportions, for those who believe in Hell and those who believe in Heaven, lies between the range 0.09422201 and 0.12542085.

Question 4

Part A

```
filename <- "http://users.stat.ufl.edu/~aa/cat/data/DeathPenalty.dat"
library(readr)
Data <- read.table(file=filename, header=T)
log.lin.mod <- glm(count ~ (D+V+P)^2, data=Data, family=poisson)
summary(log.lin.mod)

##
## Call:
## glm(formula = count ~ (D + V + P)^2, family = poisson, data = Data)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7      8
##  0.02505 -0.00895 -0.05463  0.03000 -0.60362  0.04572  0.09251 -0.01545
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.93578    0.08471  58.265  < 2e-16 ***
## Dwhite        -2.17465    0.26377  -8.245  < 2e-16 ***
## Vwhite        -1.32980    0.18479  -7.196 6.19e-13 ***
## Pyes          -3.59610    0.50691  -7.094 1.30e-12 ***
## Dwhite:Vwhite  4.59497    0.31353  14.656  < 2e-16 ***
## Dwhite:Pyes   -0.86780    0.36707  -2.364  0.0181 *
## Vwhite:Pyes    2.40444    0.60061   4.003 6.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 1225.07955  on 7  degrees of freedom
## Residual deviance:   0.37984  on 1  degrees of freedom
## AIC: 52.42
##
## Number of Fisher Scoring iterations: 3
```

The residual deviance is 0.37984 and the degrees of freedom is 1. With this information, we can compute a deviance goodness-of-fit test.

```
1-pchisq(0.37984, 1)
```

```
## [1] 0.5376889
```

H0: model fits adequately

Ha: model does not fit adequately

We have a residual deviance of 0.37984 and the degrees of freedom is 1. As a result, we get the p-value of 0.5376889. Given that the p-value is above 0.05, we can say that the model fits adequately.

Part B

```
exp(-0.8678)
```

```
## [1] 0.4198743
```

The estimated conditional odds ratio between D and P at each category of V is 0.4198743. If we were to keep the victims' race constant, then the estimated odds for a white defendant receiving the death penalty is 0.4198743 times the odds of a black defendant.

Part C

```
((53+0)/(414+16))/((11+4)/(139+37))
```

```
## [1] 1.446202
```

The marginal odds ratio between D and P is 1.446202, which means that the estimated odds of a white person receiving the death penalty is 1.446202 times the odds of a black person receiving the death penalty. Our conditional odds ratio was less than 1, but marginal odds ratio is greater than 1. This is a case of the Simpson's paradox, where a trend appears in several different groups of data but disappears or reverses when these groups are combined. If we look at the data marginally, then the black defendants have less chance of death penalty. However, if we control for victim race, then the black defendants have higher chance of death penalty.

Part D

```
V <- c("white", "white", "black", "black")
```

```
D <- c("white", "black", "white", "black")
```

```
Yes <- c(53, 11, 0, 4)
```

```
No <- c(414, 37, 16, 139)
```

```
Data2 <- data.frame(D, V, Yes, No)
```

```
Data2
```

```
##      D      V Yes  No
```

```
## 1 white white  53 414
```

```
## 2 black white   11  37
```

```
## 3 white black    0  16
```

```
## 4 black black    4 139
```

```
fit.logit <- glm(cbind(Yes,No) ~ D+V, data= Data2, family=binomial)
```

```
summary(fit.logit)
```

```
##
```

```
## Call:
```

```
## glm(formula = cbind(Yes, No) ~ D + V, family = binomial, data = Data2)
```

```
##
```

```
## Deviance Residuals:
```

```
##           1           2           3           4
## 0.02660 -0.06232 -0.60535 0.09379
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.5961      0.5069  -7.094 1.30e-12 ***
## Dwhite      -0.8678      0.3671  -2.364 0.0181 *
## Vwhite       2.4044      0.6006   4.003 6.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 22.26591 on 3 degrees of freedom
## Residual deviance: 0.37984 on 1 degrees of freedom
## AIC: 19.3
##
## Number of Fisher Scoring iterations: 4
```

Here is the corresponding model:

$$\text{logit}(\hat{\pi}) = -3.5961 - 0.8678 * \text{defendant} + 2.4044 * \text{victim}$$

As you can see, the coefficients for defendant(white) and victim(white) are the same as the coefficients for DWhite:Pyes and VWhite:PYes in the loglinear model, respectively.

Question 5

Part A

```
mbti <- read.table("http://users.stat.ufl.edu/~aa/intro-cda/data/MBTI.dat",
header = T)

mbti.mod<- glm(n ~ EI + SN + TF + JP, family = poisson, data = mbti)
summary(mbti.mod)

##
## Call:
## glm(formula = n ~ EI + SN + TF + JP, family = poisson, data = mbti)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3550  -2.1182  -1.0628   0.8506   5.7457
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.79255    0.07674  49.422 < 2e-16 ***
## EIi          0.26439    0.06226   4.246 2.17e-05 ***
## SNs          0.87008    0.06765  12.861 < 2e-16 ***
## TFt         -0.48551    0.06355  -7.640 2.17e-14 ***
## JPp         -0.12971    0.06185  -2.097 0.036 *
## ---
```



```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 399.94  on 15  degrees of freedom
## Residual deviance: 135.87  on 11  degrees of freedom
## AIC: 238.7
##
## Number of Fisher Scoring iterations: 4

### residual deviance G2
deviance(mbti.mod)

## [1] 135.8672

1 - pchisq(deviance(mbti.mod),df.residual(mbti.mod))

## [1] 0

## Pearson chisquare
sum(residuals(mbti.mod, type = "pearson")^2)

## [1] 145.1028

1 - pchisq(sum(residuals(mbti.mod, type =
"pearson")^2),df.residual(mbti.mod))

## [1] 0
```

H0: Model fits adequately. Ha: Model does not fit adequately.

The residual deviance is 135.8672 and the Pearson Chi-squared is 145.1028. In both cases, the p-value converges to zero. Given that the p-value is much below 0.05, we can reject the null hypothesis and say that the model fits poorly.

Part B

```
mbti.hmod <- glm(formula = n ~ (EI + SN + TF + JP)^2, family = poisson, data
= mbti)
summary(mbti.hmod)

##
## Call:
## glm(formula = n ~ (EI + SN + TF + JP)^2, family = poisson, data = mbti)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7
## -0.72826  1.00215  0.05168 -0.01429  1.49947 -1.29325 -0.07596
##  0.00231
##      9     10     11     12     13     14     15
##  0.56850 -0.82975 -0.04948  0.01728 -1.57051  1.09960  0.08587
##      16
```

```

0.00804
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.44760    0.13793  24.994 < 2e-16 ***
## EII         -0.02907    0.15266  -0.190 0.848952
## SNs          1.21082    0.14552   8.320 < 2e-16 ***
## Tft         -0.64194    0.16768  -3.828 0.000129 ***
## Jp           0.93417    0.14594   6.401 1.54e-10 ***
## EII:SNs      0.30212    0.14233   2.123 0.033780 *
## EII:Tft      0.19449    0.13121   1.482 0.138258
## EII:Jp       0.01766    0.13160   0.134 0.893261
## SNs:Tft      0.40920    0.15243   2.684 0.007265 **
## SNs:Jp      -1.22153    0.14547  -8.397 < 2e-16 ***
## Tft:Jp      -0.55936    0.13512  -4.140 3.48e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##    Null deviance: 399.944  on 15  degrees of freedom
## Residual deviance: 10.162  on  5  degrees of freedom
## AIC: 125
##
## Number of Fisher Scoring iterations: 4

### residual deviance G2
deviance(mbti.hmod)

## [1] 10.16171

1 - pchisq(deviance(mbti.hmod),df.residual(mbti.hmod))

## [1] 0.0707809

### Pearson chi-square
sum(residuals(mbti.hmod, type = "pearson")^2)

## [1] 10.10336

1 - pchisq(sum(residuals(mbti.hmod, type =
"pearson")^2),df.residual(mbti.hmod))

## [1] 0.07235899

```

H0: Model fits adequately. Ha: Model does not fit adequately.

The residual deviance is 10.16171 and the Pearson Chi-squared is 10.10336. The residual deviance p-value is 0.0707809 and the pearson chi-squared p-value is 0.07235899. Hence, we cannot reject the null hypothesis. We have enough evidence to conclude that this model is adequate and has a much better fit than the previous one.

Part i

The parameter estimate for the conditional log odds ratio between the S/N and J/P scales is -1.222, larger than any of the others in absolute value. Likewise, it has the smallest p-value among the conditional associations. Therefore, the estimated conditional association is strongest between the S/N and J/P scales.

Part ii

```
#wald chi square statistics
```

```
1.482 ^ 2
```

```
## [1] 2.196324
```

```
0.134 ^ 2
```

```
## [1] 0.017956
```

The corresponding Wald chi-squared statistic for the log odds ratio between the EI and TF scale is 2.20. The corresponding Wald chi-squared statistic for the log odds ratio between the EI and JP scale is 0.017956. They both have a degree of freedom equal to 1. According to this information, they have respectively the p-values of 0.138258 and 0.893261. Both of these p-values are greater than 0.05, which accordingly shows that the conditional association between the scales are insignificant.

Part C

```
mbti2.mod <- glm(n ~ EI + SN + TF + JP + EI:SN + SN:TF + SN:JP + TF:JP,  
family = poisson, data = mbti)  
summary(mbti2.mod)
```

```
##
```

```
## Call:
```

```
## glm(formula = n ~ EI + SN + TF + JP + EI:SN + SN:TF + SN:JP +  
## TF:JP, family = poisson, data = mbti)
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -1.65487 -0.46916  0.00529  0.54208  1.47431
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept)  3.41362    0.12930  26.402  < 2e-16 ***  
## EIi          0.03871    0.11361   0.341  0.733287  
## SNs          1.19414    0.14548   8.208  2.24e-16 ***  
## TFt         -0.54137    0.15282  -3.543  0.000396 ***  
## JPp          0.94292    0.13064   7.218  5.28e-13 ***  
## EIi:SNs      0.32190    0.13598   2.367  0.017922 *  
## SNs:TFt      0.42366    0.15200   2.787  0.005318 **  
## SNs:JPp     -1.22021    0.14513  -8.408  < 2e-16 ***  
## TFt:JPp     -0.55853    0.13497  -4.138  3.50e-05 ***  
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 399.944  on 15  degrees of freedom
## Residual deviance: 12.369  on  7  degrees of freedom
## AIC: 123.2
##
## Number of Fisher Scoring iterations: 4

anova(mbti2.mod, mbti.hmod, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: n ~ EI + SN + TF + JP + EI:SN + SN:TF + SN:JP + TF:JP
## Model 2: n ~ (EI + SN + TF + JP)^2
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1          7      12.369
## 2          5      10.162  2      2.207   0.3317
```

The difference in deviances is $12.369 - 10.162 = 2.207$, and the df is $7 - 5 = 2$. Therefore, we get a p-value of 0.3317. Given that this p-value is much higher than 0.05, we can say that the simpler model that assumes conditional independence between E/I and T/F, and between E/I and J/P, has a better fit than the model of homogenous association.

```
exp(confint(mbti2.mod, method="profile"))

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept) 23.4067710 38.8667659
## EIi         0.8319067  1.2992469
## SNs         2.4923799  4.4102180
## TFt         0.4299779  0.7832723
## Jp          1.9947113  3.3306215
## EIi:SNs     1.0568719  1.8015497
## SNs:TFt     1.1363105  2.0630723
## SNs:JPp     0.2214587  0.3913279
## TFt:JPp     0.4385878  0.7446507
```

The 95% likelihood-ratio confidence interval for the conditional odds ratio between the S/N and J/P scales is [0.2214587, 0.3913279]. We are 95% confident that the odds of being N given the odds of being J are between 0.2214587 and 0.3913279.

Part D

```
final.mod <- glm(formula = n ~ (EI + SN + TF + JP)^3, family = poisson, data
= mbti)
summary(mbti.mod)

##
## Call:
```

```
## glm(formula = n ~ EI + SN + TF + JP, family = poisson, data = mbti)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -4.3550   -2.1182   -1.0628    0.8506    5.7457
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.79255     0.07674  49.422 < 2e-16 ***
## EII          0.26439     0.06226   4.246 2.17e-05 ***
## SNs          0.87008     0.06765  12.861 < 2e-16 ***
## TFt         -0.48551     0.06355  -7.640 2.17e-14 ***
## JPP         -0.12971     0.06185  -2.097  0.036 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 399.94  on 15  degrees of freedom
## Residual deviance: 135.87  on 11  degrees of freedom
## AIC: 238.7
##
## Number of Fisher Scoring iterations: 4

summary(mbti.hmod)

##
## Call:
## glm(formula = n ~ (EI + SN + TF + JP)^2, family = poisson, data = mbti)
##
## Deviance Residuals:
##      1        2        3        4        5        6        7
## -0.72826  1.00215  0.05168 -0.01429  1.49947 -1.29325 -0.07596
## 0.00231
##      9       10       11       12       13       14       15
##  0.56850 -0.82975 -0.04948  0.01728 -1.57051  1.09960  0.08587 -
## 0.00804
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.44760     0.13793  24.994 < 2e-16 ***
## EII         -0.02907     0.15266  -0.190 0.848952
## SNs          1.21082     0.14552   8.320 < 2e-16 ***
## TFt         -0.64194     0.16768  -3.828 0.000129 ***
## JPP          0.93417     0.14594   6.401 1.54e-10 ***
## EII:SNs      0.30212     0.14233   2.123 0.033780 *
## EII:TFt      0.19449     0.13121   1.482 0.138258
## EII:JPP      0.01766     0.13160   0.134 0.893261
```

```

## SNs:Tft      0.40920      0.15243      2.684 0.007265 **
## SNs:JPp      -1.22153      0.14547     -8.397 < 2e-16 ***
## Tft:JPp      -0.55936      0.13512     -4.140 3.48e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 399.944  on 15  degrees of freedom
## Residual deviance: 10.162  on  5  degrees of freedom
## AIC: 125
##
## Number of Fisher Scoring iterations: 4

summary(final.mod)

##
## Call:
## glm(formula = n ~ (EI + SN + TF + JP)^3, family = poisson, data = mbti)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7       8
## -0.4805  0.6865  0.4228 -0.4746  0.9580 -0.9412 -0.7382  0.4889
##      9     10     11     12     13     14     15     16
##  0.3666 -0.5798 -0.3618  0.4228 -1.0803  0.7577  0.8099 -0.4746
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.56370     0.16197  22.002 < 2e-16 ***
## EIi          -0.27880     0.23337   -1.195  0.2322
## SNs           1.05839     0.18535   5.710 1.13e-08 ***
## Tft          -0.63483     0.25356   -2.504  0.0123 *
## JPp           0.76316     0.19243   3.966 7.31e-05 ***
## EIi:SNs       0.61460     0.25451   2.415  0.0157 *
## EIi:Tft       0.20026     0.30833   0.650  0.5160
## EIi:JPp       0.37430     0.26332   1.421  0.1552
## SNs:Tft       0.41081     0.27510   1.493  0.1353
## SNs:JPp      -0.96288     0.22994  -4.187 2.82e-05 ***
## Tft:JPp      -0.58773     0.29782  -1.973  0.0484 *
## EIi:SNs:Tft  -0.02364     0.30704  -0.077  0.9386
## EIi:SNs:JPp  -0.51039     0.29275  -1.743  0.0813 .
## EIi:Tft:JPp  0.02440     0.27403   0.089  0.9290
## SNs:Tft:JPp  0.01922     0.30880   0.062  0.9504
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 399.9439  on 15  degrees of freedom
## Residual deviance:   7.0963  on  1  degrees of freedom

```

```
## AIC: 129.93
##
## Number of Fisher Scoring iterations: 4
```

The loglinear model of mutual independence will have a total of 5 variables (1 intercept + 4 main effects).

The model of homogeneous association will have a total of 11 variables (1 intercept + 4 main effects + 6 two-factor association terms).

The model containing all three-factor interaction terms will have a total of 15 variables (1 intercept + 4 main effects + 6 two-factor association terms + 4 three-factor interaction terms).

The R output above confirms this, as well.