

# Exercise for Lecture 13

Rei Monden

## Exercise 13-1

a.

```
death.df <- read.table("Datasets/DeathPenalty.dat", header = TRUE)

death.fit <- glm(count ~ D*V + D*P + V*P,
                 family = poisson(link = "log"),
                 data = death.df)
summary(death.fit)

##
## Call:
## glm(formula = count ~ D * V + D * P + V * P, family = poisson(link = "log"),
##      data = death.df)
##
## Coefficients:
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept)   4.93578    0.08471  58.265 < 0.0000000000000002 ***
## Dwhite        -2.17465    0.26377  -8.245 < 0.0000000000000002 ***
## Vwhite        -1.32980    0.18479  -7.196  0.0000000000000619 ***
## Pyes          -3.59610    0.50691  -7.094  0.0000000000001301 ***
## Dwhite:Vwhite  4.59497    0.31353  14.656 < 0.0000000000000002 ***
## Dwhite:Pyes   -0.86780    0.36707  -2.364      0.0181 *
## Vwhite:Pyes    2.40444    0.60061   4.003  0.000062454636499 ***
## ---
## 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
```

- b. Because the model does not include the interaction effect between  $D$ ,  $P$ , and  $V$ , the odds ratio between  $D$  and  $P$  are the same at each of the two  $V$  levels. This then boils down to how to interpret a two-way interaction effect, which we learned in the lecture (slides 31-32):

*The odds ratio between  $D$  and  $P$  are equal to  $\exp(-0.8678) = .42$ .*

Thus, for instance, regardless of the victim's race, the odds of a white over a black defendant for those receiving the death penalty is 0.42 times the odds of a white over a black defendant for those not receiving the death penalty. In other words, receiving the death penalty is less likely for white defendants in comparison to black defendants.

- c. From the output in a., we see that the residual deviance is 0.380 with 1 df. The likelihood ratio test comparing this model to the saturated model is therefore  $\chi(1) = 0.380$ ,  $p = 0.54$ . We conclude that our model does not fit significantly worse than the saturated model.

We could also fit the saturated model and compare this to our model explicitly:

```
death.sat <- glm(count ~ D*V*P,
                 family = poisson(link = "log"),
                 data = death.df)
anova(death.fit, death.sat, test = "LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: count ~ D * V + D * P + V * P
## Model 2: count ~ D * V * P
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1          1    0.37984
## 2          0    0.00000  1  0.37984  0.5377
```

- d. The estimated loglinear model from a. is

$$\log(\mu_{ijk}) = 4.936 - 2.175D_{\text{white}} - 1.330V_{\text{white}} - 3.596P_{\text{yes}} \\ + 4.595(D_{\text{white}} \times V_{\text{white}}) - 0.868(D_{\text{white}} \times P_{\text{yes}}) + 2.404(V_{\text{white}} \times P_{\text{yes}}),$$

where  $D_{\text{white}}$ ,  $V_{\text{white}}$ , and  $P_{\text{yes}}$  are 0-1 indicator variables. The desired predicted count is

$$\exp[4.936 - 2.175(1) - 1.330(0) - 3.596(0) + 4.595(0) - 0.868(0) + 2.404(0)] = 15.82.$$

The observed count is 16, thus the predicted count is very close.

## Exercise 13-2

a.

```
MBTI.df <- read.table("Datasets/MBTI.dat", header = TRUE)

MBTI.fit <- glm(n ~ EI + SN + TF + JP,
               family = poisson(link = "log"),
               data = MBTI.df)
summary(MBTI.fit)

##
## Call:
## glm(formula = n ~ EI + SN + TF + JP, family = poisson(link = "log"),
##      data = MBTI.df)
##
## Coefficients:
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept)  3.79255    0.07674  49.422 < 0.0000000000000002 ***
## EIi          0.26439    0.06226   4.246  0.0000217200004913 ***
## SNs          0.87008    0.06765  12.861 < 0.0000000000000002 ***
## TFt         -0.48551    0.06355  -7.640  0.00000000000000217 ***
## 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
```

- b. From the output in a., we see that the residual deviance is 135.87 with 11 df. The likelihood ratio test comparing this model to the saturated model is therefore  $\chi(11) = 135.87$ ,  $p < .001$ . We conclude that our model fits significantly worse than the saturated model.

We could also fit the saturated model and compare this to our model explicitly:

```
MBTI.sat <- glm(n ~ EI * SN * TF * JP,
               family = poisson(link = "log"),
               data = MBTI.df)
anova(MBTI.fit, MBTI.sat, test = "LRT")

## Analysis of Deviance Table
##
## Model 1: n ~ EI + SN + TF + JP
## Model 2: n ~ EI * SN * TF * JP
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1 11 135.87
## 2 0 0.00 11 135.87 < 0.00000000000000022 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c.

```
MBTI.fit2 <- glm(n ~ (EI + SN + TF + JP)^2,
                family = poisson(link = "log"),
                data = MBTI.df)
summary(MBTI.fit2)

##
## Call:
## glm(formula = n ~ (EI + SN + TF + JP)^2, family = poisson(link = "log"),
## data = MBTI.df)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 3.44760 0.13793 24.994 < 0.0000000000000002 ***
## EIi -0.02907 0.15266 -0.190 0.848952
## SNs 1.21082 0.14552 8.320 < 0.0000000000000002 ***
## TFt -0.64194 0.16768 -3.828 0.000129 ***
## JPp 0.93417 0.14594 6.401 0.000000000154 ***
## 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 < 0.0000000000000002 ***
## TFt:JPp -0.55936 0.13512 -4.140 0.000034776300 ***
## ---
## 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
```

As can be seen, the strongest interaction effect is between S/N and J/P (−1.222).

- d. From the output in c., it can be seen that both the conditional association between the E/I and T/F scales (0.194) and the conditional association between the E/I and J/P scales (0.0018) are not statistically significant at the 5% significance level. Thus, in both cases, we fail to reject the null hypothesis of no conditional association.

### Exercise 13-3

a.

```
MBTI.fit3 <- glm(n ~ EI + SN + TF + JP + EI:SN + SN:TF + SN:JP + TF:JP,
                 family = poisson(link = "log"),
                 data = MBTI.df)
summary(MBTI.fit3)
```

```
##
## Call:
## glm(formula = n ~ EI + SN + TF + JP + EI:SN + SN:TF + SN:JP +
##      TF:JP, family = poisson(link = "log"), data = MBTI.df)
##
## Coefficients:
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept)  3.41362    0.12930  26.402 < 0.0000000000000002 ***
## EIi          0.03871    0.11361   0.341      0.733287
## SNs          1.19414    0.14548   8.208 0.000000000000000224 ***
## TFt         -0.54137    0.15282  -3.543    0.000396 ***
## JPs          0.94292    0.13064   7.218 0.00000000000000528411 ***
## 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 < 0.0000000000000002 ***
## TFt:JPp     -0.55853    0.13497  -4.138 0.000035027315884390 ***
## ---
## 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
```

b.

```
anova(MBTI.fit3, MBTI.fit2, test = "LRT")
```

```
## 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 result is  $\chi^2(2) = 2.21$ ,  $p = .33$ , thus there is no significant difference between this model and the model containing all pairwise associations.

c.

```
confint(MBTI.fit3)
```

```
## Waiting for profiling to be done...
```

```
##           2.5 %    97.5 %
## (Intercept) 3.1530253 3.6601395
## EIi        -0.1840350 0.2617848
## SNs         0.9132380 1.4839241
## TFt        -0.8440215 -0.2442749
## JPP         0.6904993 1.2031589
## EIi:SNs     0.0553135 0.5886472
## SNs:TFt     0.1277866 0.7241963
## SNs:JPP    -1.5075192 -0.9382094
## TFt:JPP    -0.8241952 -0.2948400
```

From the output we see that the required 95% confidence interval is  $(\exp(-1.508), \exp(-0.938)) = (0.22, 0.39)$ .