Hw9

Mohammad Mottahedi

March 28, 2016

```
set.seed(1234)
library(MASS)
```

Problem 1

part a

```
imp.treat.a <- c(8, 7, 6, 6, 3, 4, 7, 2, 3, 4)
imp.treat.b <- c(9, 9,8, 14, 8, 13, 11, 5, 7, 7)
df <- data.frame(imperfection = c(imp.treat.a,imp.treat.b),
                 trearment = c(rep(1,10), rep(0,10)))
model <- glm(imperfection ~ 1 + trearment, data = df,
             family = poisson(link = log))
summary(model)
```

```
##
## Call:
## glm(formula = imperfection ~ 1 + trearment, family = poisson(link = log),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5280  -0.7259  -0.2028   0.6680   1.5040
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.2083     0.1048  21.066  < 2e-16 ***
## trearment    -0.5988     0.1760  -3.402  0.00067 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 27.700  on 19  degrees of freedom
## Residual deviance: 15.604  on 18  degrees of freedom
## AIC: 93.835
##
## Number of Fisher Scoring iterations: 4
```

$log(\mu_A) = 1.6094379,$

$log(\mu_B) = 2.2082744$

$\beta = log(\mu_A) - log(\mu_B) = -0.5988365$

interpretation: The expected number of imperfection is $e^{\beta} = 0.7624804$ times smaller when we do treatment A.

part b

$\mu_A = \beta_0 + \beta_1, \mu_B = \beta_0$

$\mu_A - \mu_B = \beta_1$

$H_0 = \beta_1 = \mu_A - \mu_B = 0$

$H_0 = \mu_A = \mu_B$

$H_A = \mu_A \neq \mu_B$

$z^2 = (-0.5988/0.176)^2 = 11.5754597$

p-value= $6.6827912 \times 10^{-4} < 0.05$

so we reject the null hypothesis that expected numbers are the same for both treatment and treatment type is significant.

Problem 2

part a

```
hv <- data.frame(age = c("<55","<55",">55",">55"),
                 type = c("aortic","mitral","aortic","mitral"),
                 exposure = c(1259,2082,1417,1647),
                 deaths = c(4,1,7,9))
lexposure = log(hv$exposure)
model <- glm(deaths ~ age + type + type*age, offset = lexposure, data = hv, family = poisson(link = log
summary(model)
```

```
##
## Call:
## glm(formula = deaths ~ age + type + type * age, family = poisson(link = log),
##     data = hv, offset = lexposure)
##
## Deviance Residuals:
## [1]  0  0  0  0
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -5.7518     0.5000 -11.504   <2e-16 ***
## age>55              0.4414     0.6268   0.704   0.4813
## typemitral         -1.8893     1.1180  -1.690   0.0911 .
## age>55:typemitral   1.9902     1.2264   1.623   0.1046
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
```

```
##     Null deviance: 1.0841e+01  on 3  degrees of freedom
## Residual deviance: 6.6613e-16  on 0  degrees of freedom
## AIC: 21.127
##
## Number of Fisher Scoring iterations: 3
```

```
anova(model)
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: deaths
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                         3    10.8405
## age       1   7.0508         2     3.7897
## type      1   0.5672         1     3.2225
## age:type  1   3.2225         0     0.0000
```

do you know the estimate of this value? 4

When you take the natural log of this estimate, which parameter estimate in your model do you expect to get?

$log(\mu_1) = 1.3862944 \approx -5.7518 + log(1259) = 1.386273$

part b

with 95% confidence interval, only the intercept is significant the rest of coefficients all have P > 0.05.

Given the the valve type the estimated death rate for older group is 1.5589305 times the younger group and given the age, group with mitral valve replacement has 0.1511776 times smaller death rate.

part c

to adjust for the time interval of individual observations.

part d

```
model <- glm(deaths ~ age + type, offset = lexposure, data = hv, family = poisson(link = log) )
summary(model)
```

```
##
## Call:
## glm(formula = deaths ~ age + type, family = poisson(link = log),
```

```
##     data = hv, offset = lexposure)
##
## Deviance Residuals:
##      1       2       3       4
##  1.025  -1.197  -0.602   0.613
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.3121     0.5066 -12.460   <2e-16 ***
## age>55        1.2209     0.5138   2.376   0.0175 *
## typemitral   -0.3299     0.4382  -0.753   0.4515
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 10.8405  on 3  degrees of freedom
## Residual deviance:  3.2225  on 1  degrees of freedom
## AIC: 22.349
##
## Number of Fisher Scoring iterations: 5
```

The saturated model is to complex and it's probably over fit. The main effect model is more appropriate. also, the interaction term doesn't seem to be significant.