Contingency Analysis

Yang

April 2, 2019

Load Dataset

```r
Y <- c(76,160,6,25,114,181,11,48)
G <- factor(c("M","M","M","M","F","F","F","F"),levels=c("M","F"))
I <- factor(c("Yes","Yes","No","No","Yes","Yes","No","No"),levels=c("Yes","No"))
H <- factor(c("Yes","No","Yes","No","Yes","No","Yes","No"),levels=c("Yes","No"))

GSS <- data.frame(cbind(Y,G,I,H))
# Create a contingency table for a better view
Tab <- xtabs(Y~G+I+H)

m_full <- glm(Y~G*I*H,family=poisson)
summary(m_full)
```

```
##
## Call:
## glm(formula = Y ~ G * I * H, family = poisson)
##
## Deviance Residuals:
## [1]  0  0  0  0  0  0  0  0
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.3307     0.1147  37.754  < 2e-16 ***
## GF            0.4055     0.1481   2.738  0.00618 **
## INo          -2.5390     0.4241  -5.987 2.13e-09 ***
## HNo           0.7444     0.1393   5.344 9.11e-08 ***
## GF:INo        0.2007     0.5287   0.380  0.70427
## GF:HNo       -0.2821     0.1836  -1.537  0.12434
## INo:HNo       0.6827     0.4755   1.436  0.15106
## GF:INo:HNo    0.3283     0.5934   0.553  0.58005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4.4582e+02  on 7  degrees of freedom
## Residual deviance: 1.4433e-14  on 0  degrees of freedom
## AIC: 61.382
##
## Number of Fisher Scoring iterations: 3
```

Model Selection Through Backward Elimination

```
m2<-update(m_full,~.-G:I:H)
anova(m2,m_full,test="Chi")
```

```
## Analysis of Deviance Table
##
## Model 1: Y ~ G + I + H + G:I + G:H + I:H
## Model 2: Y ~ G * I * H
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         1    0.30072
## 2         0    0.00000  1  0.30072   0.5834
```

```
summary(m2)
```

```
##
## Call:
## glm(formula = Y ~ G + I + H + G:I + G:H + I:H, family = poisson)
##
## Deviance Residuals:
##        1         2         3         4         5         6         7
## -0.10362   0.07183   0.39073  -0.17923   0.08516  -0.06730  -0.26626
##        8
##   0.13173
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.3426     0.1120  38.763  < 2e-16 ***
## GF            0.3856     0.1434   2.689  0.00717 **
## INo          -2.7147     0.3035  -8.945  < 2e-16 ***
## HNo           0.7269     0.1353   5.374 7.68e-08 ***
## GF:INo        0.4636     0.2406   1.927  0.05401 .
## GF:HNo       -0.2516     0.1749  -1.438  0.15035
## INo:HNo       0.8997     0.2852   3.155  0.00160 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 445.82335  on 7  degrees of freedom
## Residual deviance:   0.30072  on 1  degrees of freedom
## AIC: 59.683
##
## Number of Fisher Scoring iterations: 4
```

```
m3<-update(m2,~.-G:H)
anova(m3,m2,test="Chi")
```

```
## Analysis of Deviance Table
##
## Model 1: Y ~ G + I + H + G:I + I:H
## Model 2: Y ~ G + I + H + G:I + G:H + I:H
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         2    2.38314
## 2         1    0.30072  1   2.0824    0.149
```

```r
summary(m3)
```

```
## 
## Call:
## glm(formula = Y ~ G + I + H + G:I + I:H, family = poisson)
## 
## Deviance Residuals:
##        1        2        3        4        5        6        7
## -0.93493  0.67971  0.05945 -0.02883  0.81131 -0.61817 -0.04336
##        8
##  0.02087
## 
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.43609    0.08728  50.828  < 2e-16 ***
## GF           0.22314    0.08733   2.555  0.01062 *
## INo         -2.66870    0.29595  -9.017  < 2e-16 ***
## HNo          0.58486    0.09053   6.460 1.04e-10 ***
## GF:INo       0.42041    0.23840   1.763  0.07782 .
## INo:HNo      0.87239    0.28411   3.071  0.00214 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for poisson family taken to be 1)
## 
##     Null deviance: 445.8233  on 7  degrees of freedom
## Residual deviance:   2.3831  on 2  degrees of freedom
## AIC: 59.765
## 
## Number of Fisher Scoring iterations: 3
```

```r
m4<-update(m3,~.-G:I)
anova(m4,m2,test="Chi")
```

```
## Analysis of Deviance Table
## 
## Model 1: Y ~ G + I + H + I:H
## Model 2: Y ~ G + I + H + G:I + G:H + I:H
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         3     5.5810
## 2         1     0.3007  2   5.2803  0.07135 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(m4)
```

```
## 
## Call:
## glm(formula = Y ~ G + I + H + I:H, family = poisson)
## 
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
## -0.6372   1.0893  -0.4999  -1.1823   0.5421  -0.9715   0.4116   0.9662
## 
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.40294    0.08601   51.189  < 2e-16 ***
## GF           0.28205    0.08106    3.480 0.000502 ***
## INo         -2.41381    0.25315   -9.535  < 2e-16 ***
## HNo          0.58486    0.09053    6.460 1.04e-10 ***
## INo:HNo      0.87239    0.28411    3.071 0.002136 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 445.823  on 7  degrees of freedom
## Residual deviance:   5.581  on 3  degrees of freedom
## AIC: 60.963
##
## Number of Fisher Scoring iterations: 4
```

```r
anova(m_full,m2,m3,m4,test = "Chi")
```

```
## Analysis of Deviance Table
##
## Model 1: Y ~ G * I * H
## Model 2: Y ~ G + I + H + G:I + G:H + I:H
## Model 3: Y ~ G + I + H + G:I + I:H
## Model 4: Y ~ G + I + H + I:H
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         0     0.0000
## 2         1     0.3007 -1  -0.3007  0.58343
## 3         2     2.3831 -1  -2.0824  0.14900
## 4         3     5.5810 -1  -3.1979  0.07373 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
m_full$deviance
```

```
## [1] 1.44329e-14
```

```r
m2$deviance
```

```
## [1] 0.3007213
```

```r
m3$deviance
```

```
## [1] 2.383139
```

```r
m4$deviance
```

```
## [1] 5.581035
```

Model 4 Y ~ G + I + H +I:H is selected as a adequate simplification of full model through backward elimination. And the for the interaction term(I:H),if any of I or H is given, it is conditionally independent of the rest.

Confidence Interval

```
#95% CI
m_b <- glm(Y~H*I+I*G,family=poisson)
coefs =summary(m_b)$coef
exp(confint(m_b)["HNo:INo",])
```

Waiting for profiling to be done...

```
##     2.5 %   97.5 %
## 1.402592 4.299645
```
```
#odds ratio
e<-exp(1)
e^(coefs[5,1])
```

```
## [1] 2.392617
```

Interpretation

The odds ratio is $e^{0.8723877} = 2.392646$, meaning the odds ratio of people thinking that the government pay all of the health care costs of AIDS patients and there should be a government information program to promote safe sex practices and the odds of people who disagree with the above is 2.392646. And it does not change as the gender changes since it is independent of gender.