

# LNL\_\_HW\_\_week6

*Patrick Kelly*

*Saturday, February 28, 2015*

## Week 7 Assignment: Contingency Tables

The contingency table that we created today shows counts of preferences for tea or coffee cross-classified according to whether the responder is a male or a female.

```
y<-c(8,4,5,2)
Gender<-gl(2,1,4,labels=c("Male","Female"))
Preference<-gl(2,2,labels=c("Coffee","Tea"))
Beverage.Preference<-data.frame(y,Gender,Preference)
Beverage.Preference
```

```
##   y Gender Preference
## 1 8   Male     Coffee
## 2 4 Female     Coffee
## 3 5   Male       Tea
## 4 2 Female       Tea
```

Analyze this table using material from Chapter 4 of the book and the workshop description for this week.

```
(bp.table <- xtabs(y ~ Gender+Preference))
```

```
##           Preference
## Gender   Coffee Tea
##   Male         8   5
##   Female        4   2
```

Estimate Poisson and binomial models.

```
#Poisson
p.mod1 <- glm(y ~ Gender+Preference, poisson)
summary(p.mod1)

##
## Call:
## glm(formula = y ~ Gender + Preference, family = poisson)
##
## Deviance Residuals:
##      1       2       3       4
## -0.07379  0.10717  0.09551 -0.14394
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.1054     0.3281   6.418 1.38e-10 ***
## GenderFemale  -0.7732     0.4935  -1.567   0.117
```

```
## PreferenceTea -0.5390      0.4756 -1.133      0.257
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 4.01889  on 3  degrees of freedom
## Residual deviance: 0.04677  on 1  degrees of freedom
## AIC: 19.345
##
## Number of Fisher Scoring iterations: 3
```

```
#Binomial
(m <- matrix(y,nrow=2))
```

```
##      [,1] [,2]
## [1,]    8    5
## [2,]    4    2
```

```
b.mod1 <- glm(m ~ 1, family=binomial)
```

Apply `drop1()` to analyze significance of factors. Describe how `drop1()` works.

```
drop1(p.mod1, test = "Chi")
```

```
## Single term deletions
##
## Model:
## y ~ Gender + Preference
##      Df Deviance   AIC    LRT Pr(>Chi)
## <none>      0.04677 19.345
## Gender    1  2.68748 19.986 2.6407  0.1042
## Preference 1  1.37818 18.676 1.3314  0.2486
```

How `drop1()` works: This function fits different models by dropping each of the arguments within the model and provides a comparison of models based on the AIC criterion and whatever test method specified (Chisquared in this case). Generally the lower the AIC, the better the model. It is also important to consider the `Pr(>Chi)` terms to understand if significance of the deviance measure for each model.

In analyzing the significance of the factors, we see that both of the factors are not highly significant but are still directional. **Gender** is more significant but still not significant at a 95% confidence level, while **Preference** is loosely directional, if at all. Although the AIC increases using **Gender** and excluding **Preference**, this analysis does not make a very strong case for either parameter. This is also supported by the coefficient p-values from summary output of the poisson model.

### Analyze the fit using deviance.

In observing the deviance of the null model (4.02 on 3 degrees of freedom) we see that the null model is generally a satisfactory fit. Although the deviance when including both **Gender** and **Preference** drops to 0.05, the cost of adding adding these parameters to the model may not be worth the drop in deviance given the acceptable level of the null model.

In checking the output of the `drop1()` function, we see that including only **Preference** drops the deviance to 2.69 and including only **Gender** drops the deviance to 1.38. However considering the weak strength of

these variables on the model, the above conclusion holds that the null model is a satisfactory fit. Although there could be a case made for including **Gender**.

#### **Fit binomial model to test homogeneity.**

```
summary(b.mod1)
```

```
##
## Call:
## glm(formula = m ~ 1, family = binomial)
##
## Deviance Residuals:
##      1      2
## -0.1207  0.1795
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.5390     0.4756   1.133   0.257
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 0.04677  on 1  degrees of freedom
## Residual deviance: 0.04677  on 1  degrees of freedom
## AIC: 7.2719
##
## Number of Fisher Scoring iterations: 3
```

In observing the results of the fit of the null binomial model, we see that this hypothesis of homogeneity is supported. The deviance (0.5 on 1 degree of freedom) definitely supports the conclusions drawn above that the null model is sufficient.