Week 7 Homework

Brian Ritz

The contingency table that we created Saturday 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
```

We see just by eyeing the table that there are more males than females, and there is an overall preference for coffee over tea, but there does not appear to be an interaction.

Estimate a poisson model

```
summary(poisson.model<-glm(y~Gender+Preference, Beverage.Preference, family=poisson ))
```

```
##
```

```
## Call:
## glm(formula = y ~ Gender + Preference, family = poisson, data = Beverage.Preference)
##
## 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
```

The null model (the hypothesis that all four observations happen at the same rate) may fit the model reasonably well because the deviance of the model is 4.01 with 3 degress of freedom. Neither predictor is significant at the .1 p-value, although gender is close.

```
# apply drop 1 to one of the factors
```

```
drop1(poisson.model)
```

```
## Single term deletions
##
## Model:
## y ~ Gender + Preference
##            Df Deviance    AIC
## <none>          0.04677 19.345
## Gender      1   2.68748 19.986
## Preference  1   1.37818 18.676
```

Drop1() tests the significance of factors by dropping it from the model, and comparing the AIC from the full model with the model without the dropped term. We see our output that the model is just slightly better without the Preference term.

Estimate a binomial model Binomial model I create 2 groups and check if they have the same probability of success.

```
(binomial.matrix<-matrix(y,nrow=2))
```

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

```
binomial.model<-glm(binomial.matrix~1,family=binomial)
summary(binomial.model)
```

```
##
## Call:
```

```
## glm(formula = binomial.matrix ~ 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
```

The binomial model tests the hypothesis that the there is no effect of male/female on the coffee/tea decision.

To test for homogeneity I also fit the null model (the numll model suggests that the probability of being male is the same whether you like coffee or tea). The null deviance is similar to the residual deviance in this case, suggesting that there is homogeneity in the dataset.