Lab 8 Part 1

Dr. Isabel Sassoon

Table Analysis

In this lab we will explore how to perform table analysis. This is useful when there are multiple columns of data that are categorical.

```
# installing a new library
library(vcdExtra)
```

```
## Loading required package: vcd
```

```
## Loading required package: grid
```

```
## Loading required package: gnm
```

The data we are using is called Arthritis

```
help(Arthritis)
```

Lets explore the data

```
summary(Arthritis)
```

```
##        ID          Treatment      Sex          Age          Improved
##  Min.   : 1.00   Placebo:43   Female:59   Min.   :23.00   None  :42
##  1st Qu.:21.75   Treated:41   Male  :25   1st Qu.:46.00   Some  :14
##  Median :42.50                            Median :57.00   Marked:28
##  Mean   :42.50                            Mean   :53.36
##  3rd Qu.:63.25                            3rd Qu.:63.00
##  Max.   :84.00                            Max.   :74.00
```

We can see there is an ID, Gender, Age and whether the person had a Marked Improvement, Some improvement or none.

Table analysis can help us if we want to find out if there is a relationship between the improvement and the treatment (for example).

```
table(Arthritis$Improved, Arthritis$Treatment)
```

```
##
##          Placebo Treated
##   None        29      13
##   Some         7       7
##   Marked       7      21
```

Null hypothesis: H_0, The treatment and improvement are independent ***Alternative hypothesis***: H_1, there is a relationship between the treatment and the improvement

```
chisq.test(table(Arthritis$Improved, Arthritis$Treatment))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(Arthritis$Improved, Arthritis$Treatment)
## X-squared = 13.055, df = 2, p-value = 0.001463
```

We can see that our χ^2 is significant - so there is a dependence between treatment and outcome.

Note that there is no cell in the table that is so small that we need to consider using Fisher's exact test.

We can now take this one step further and see if we can answer the question: What attributes affect the success of the treatment?

If we wanted to model this relationship further we could transform the dependent variable from one with three values to one with two and then use Logistic regression:

Let's define the new dependent variable:

- 0 if there is none
- 1. When there is an improvement (some or marked)

```
Arthritis$Improved.Ind<-ifelse(Arthritis$Improved=="None", 0,1)
```

Now we can model this:

```
Arthritis.lr<-glm(Arthritis$Improved.Ind~Arthritis$Treatment, family = "binomial")
summary(Arthritis.lr)
```

```
##
## Call:
## glm(formula = Arthritis$Improved.Ind ~ Arthritis$Treatment, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.51567  -0.88759  -0.00712   0.87335   1.49809
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -0.7282     0.3254  -2.238  0.02524 *
## Arthritis$TreatmentTreated    1.4955     0.4675   3.199  0.00138 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 116.45  on 83  degrees of freedom
## Residual deviance: 105.49  on 82  degrees of freedom
## AIC: 109.49
##
## Number of Fisher Scoring iterations: 4
```

This confirms our table analysis. We can also look at the coefficients as Odds Ratios

```
exp(coef(Arthritis.lr))
```

```
##                (Intercept) Arthritis$TreatmentTreated
##                  0.4827586                  4.4615385
```

But what about the other possible covariates?

```
Arthritis2.lr<-glm(Arthritis$Improved.Ind~Arthritis$Treatment+ Arthritis$Sex+ Arthritis$Age, family = "
summary(Arthritis2.lr)
```

```
##
## Call:
## glm(formula = Arthritis$Improved.Ind ~ Arthritis$Treatment +
##     Arthritis$Sex + Arthritis$Age, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.10833  -0.91158   0.05362   0.91681   1.84659
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -3.01546    1.16777  -2.582  0.00982 **
## Arthritis$TreatmentTreated  1.75980    0.53650   3.280  0.00104 **
## Arthritis$SexMale          -1.48783    0.59477  -2.502  0.01237 *
## Arthritis$Age               0.04875    0.02066   2.359  0.01832 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 116.449  on 83  degrees of freedom
## Residual deviance:  92.063  on 80  degrees of freedom
## AIC: 100.06
##
## Number of Fisher Scoring iterations: 4
```

```
exp(coef(Arthritis2.lr))
```

```
##                (Intercept) Arthritis$TreatmentTreated
##                 0.04902324                 5.81129902
##          Arthritis$SexMale               Arthritis$Age
##                 0.22586198                 1.04995419
```

We can also start with the interactions model

```r
Arthritis3.lr<-glm(Arthritis$Improved.Ind~Arthritis$Treatment*Arthritis$Sex*Arthritis$Age, family = "bi
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(Arthritis3.lr)
```

```
##
## Call:
## glm(formula = Arthritis$Improved.Ind ~ Arthritis$Treatment *
##     Arthritis$Sex * Arthritis$Age, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.39213  -0.90975   0.00018   0.88692   1.95478
##
## Coefficients:
##                                                                  Estimate
## (Intercept)                                                    -3.695e+00
## Arthritis$TreatmentTreated                                     -1.318e-01
## Arthritis$SexMale                                              -2.086e+03
## Arthritis$Age                                                   6.273e-02
## Arthritis$TreatmentTreated:Arthritis$SexMale                    2.090e+03
## Arthritis$TreatmentTreated:Arthritis$Age                        3.334e-02
## Arthritis$SexMale:Arthritis$Age                                 3.336e+01
## Arthritis$TreatmentTreated:Arthritis$SexMale:Arthritis$Age     -3.347e+01
##                                                                Std. Error z value
## (Intercept)                                                     1.885e+00  -1.960
## Arthritis$TreatmentTreated                                      3.022e+00  -0.044
## Arthritis$SexMale                                               1.933e+05  -0.011
## Arthritis$Age                                                   3.407e-02   1.841
## Arthritis$TreatmentTreated:Arthritis$SexMale                    1.933e+05   0.011
## Arthritis$TreatmentTreated:Arthritis$Age                        5.644e-02   0.591
## Arthritis$SexMale:Arthritis$Age                                 3.097e+03   0.011
## Arthritis$TreatmentTreated:Arthritis$SexMale:Arthritis$Age      3.097e+03  -0.011
##                                                                Pr(>|z|)
## (Intercept)                                                      0.0500 .
## Arthritis$TreatmentTreated                                       0.9652
## Arthritis$SexMale                                                0.9914
## Arthritis$Age                                                    0.0656 .
## Arthritis$TreatmentTreated:Arthritis$SexMale                     0.9914
## Arthritis$TreatmentTreated:Arthritis$Age                         0.5547
## Arthritis$SexMale:Arthritis$Age                                  0.9914
## Arthritis$TreatmentTreated:Arthritis$SexMale:Arthritis$Age       0.9914
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 116.45  on 83  degrees of freedom
## Residual deviance:  81.18  on 76  degrees of freedom
## AIC: 97.18
##
## Number of Fisher Scoring iterations: 21
```

Note that all the interactions are not significant. The AIC is 97.18, which is very close to the one obtained without the interactions (100.06).