R Demonstration – Categorical Analysis

Objective: The purpose of this week's session is to demonstrate how to perform a categorical analysis in R. In the first part, we will examine two way contingency tables. In the second part, we will investigate multi-way contingency tables.

Part I. Two-way contingency tables

NOTE: This part of the exercise assumes that you have downloaded the dataset from Sinclair & Arcese (1995; Ecology 76: 882-891) and saved it in your PCB6466 folder as a tab-delimited text file named *sinclair.txt*. You also need to download the *Categorical_analysis.R* script and save it in your PCB6466 folder.

After starting R, change the directory to your `PCB6466` folder and open the *Categorical_analysis.R* script. The first two lines of the script read and attach the Sinclair & Arcese (1995) dataset:

```
## read the data from file and attach it
count_data <- read.table("sinclair.txt", header=T)
attach(count_data)
```

Next, we use `xtabs` function to create two way contingency tables for each two-way combination of variables: death, marrow, and sex, we use the `mosaicplot` function to plot them, and the function `chisq.test` to perform a Pearson's chi-squared test on the 2-way tables:

```
### PART I. 2-Way Contingency Tables ###

## use the xtabs() function to create 2-way contingency tables
marrow_death_xtab <- xtabs(count ~ death + marrow)
sex_death_xtab <- xtabs(count ~ death + sex)
marrow_sex_xtab <- xtabs(count ~ sex + marrow)

## use the mosaicplot() function to display the 2-way tables
par(mfrow=c(1,3))
mosaicplot(marrow_death_xtab, shade=T, main="Marrow vs. Death")
mosaicplot(sex_death_xtab, shade=T, main="Sex vs. Death")
mosaicplot(marrow_sex_xtab, shade=T, main="Marrow vs. Sex")

par(mfrow=c(1,1))

## perform Pearson's chi-squared test on the 2-way tables
chisq.test(marrow_death_xtab, correct=F)
chisq.test(sex_death_xtab, correct=F)
chisq.test(marrow_sex_xtab, correct=F)
```

From the mosaic plots and the tests we see that death is not-independent of marrow type ($P < 0.001$), but is independent of sex ($P = 0.09$), and sex is independent of cause of death ($P = 0.77$). Notice fewer carcasses than expected dead after predation when they have TG

marrow (indicating a healthy animal) and more dead after other causes (orange and red colors in the plot).

Part II. Multi-way contingency tables

Now we analyze all the data using multi-way tables. We start creating an empty table to hold the results (slide 39 of the Powerpoint)…

```
### PART II. Log-Linear Models for Multi-Way Contingency Tables ###

## create an empty table to hold the results (see Slide 39 of
Powerpoint)
colnames <- c("Model","G-squared","df","P-value","AIC")
rownames <- seq(1,9)
gtable <- matrix(nrow=9,ncol=5,dimnames=list(rownames, colnames))
```

Next we create hierarchical log-linear models. Because these data involves counts we use a Poisson error. We fit log-linear models with different combinations of terms. The fit of each model is based on comparing observed and fitted cell frequencies and, equivalently, comparing the fit of each model to that of the saturated model with zero degrees of freedom.

```
## create hierarchical log-linear models
model1 <- glm(count ~ death + sex + marrow, family=poisson)
model2 <- glm(count ~ death + sex + marrow + death:sex,
family=poisson)
model3 <- glm(count ~ death + sex + marrow + death:marrow,
family=poisson)
model4 <- glm(count ~ death + sex + marrow + sex:marrow,
family=poisson)
model5 <- glm(count ~ death*sex + death*marrow, family=poisson)
model6 <- glm(count ~ death*sex + sex*marrow, family=poisson)
model7 <- glm(count ~ death*marrow + sex*marrow, family=poisson)
model8 <- glm(count ~ death*sex + death*marrow + sex*marrow,
family=poisson)
model9 <- glm(count ~ death*sex*marrow, family=poisson)

## combine all of the models into a list and then loop over each one
models <-
list(model1,model2,model3,model4,model5,model6,model7,model8,model9)
for (i in 1:length(models)) {
## write the formula, G-squared, df, P-value, and AIC for the model
to gtable
      model <- models[[i]]
      gtable[i,1] <- toString(model$formula)
      gtable[i,2] <- model$deviance
      gtable[i,3] <- model$df.residual
      gtable[i,4] <- 1-pchisq(model$deviance, model$df.residual)
      gtable[i,5] <- model$deviance - 2*model$df.residual
}

## display the table with the model results
gtable[9,5] <- ""       ## there is no AIC for the saturated model
gtable
```

For hypothesis testing, we would fit these models hierarchically, starting with the most complex. The AIC chose model 7 as best fit, whereas G^2 chose model 8. The comparison of the fit of model 8 and the saturated model 9 is a test of the Ho that there is no three way interaction. The G^2 deviance statistic results in rejection of this Ho.

```
## test for complete independence
anova(model1, model8, test="Chi")
## test for three-way interaction
anova(model8, model9, test="Chi")
```

We also illustrate the tests for conditional independence and complete independence, although the presence of a three way interaction would usually preclude tests of two way interactions and the presence of both complete and conditional dependence would preclude testing for complete independence

```
## tests for conditional independence
anova(model7, model8, test="Chi")   ## test death vs. sex
anova(model6, model8, test="Chi")   ## test death vs. marrow
anova(model5, model8, test="Chi")   ## test sex vs. marrow
```

This demonstrates that we would reject the Ho of conditional independence of cause of death and marrow type. The three way interaction is shown by the complex patterns of odds ratios that is different for males and females. The odds of being killed by predation were less for male wildebeest with either TG marrow than OG or SWF marrow. For females, the odds of being killed by non-predator causes were less for those with SWF marrow than OG or TG. This is a different conclusion compared to that we could arrive if we stop at the two-way tables (shown below).

```
## tests for marginal independence
anova(model4, model1, test="Chi")
anova(model3, model1, test="Chi")
anova(model2, model1, test="Chi")
```

Finally we plot to show the results of the multi-way analysis.

```
## create a contingency table of death vs. marrow for females
female_data <- count_data[sex=="FEMALE",]
female_xtab <- xtabs(count ~ death + marrow, data=female_data)

## create a contingency table of death vs. marrow for males
male_data <- count_data[sex=="MALE",]
male_xtab <- xtabs(count ~ death + marrow, data=male_data)
```

```
## plot the contingency tables
par(mfrow=c(1,2))
mosaicplot(female_xtab, shade=T, main="Females")
mosaicplot(male_xtab, shade=T, main="Males")
par(mfrow=c(1,1))

## detach the data
detach(count_data)
```