Specialist Services Gender Ratio

## Aim

Determine whether the gender ratio of staff in specialist services, by grade, is significantly different from a ratio of one to one.

## The data

Provided by Corrie:

grade <- as.factor(c(1, 2, 3, 4))  
role <- c("Chief Scientist", "Principal Specialist", "Senior Specialist", "Specialist")  
female <- c(0, 4, 46, 18)  
male <- c(1, 14, 42, 10)  
  
grades <- data.frame(grade, role, female, male)  
grades

## grade role female male  
## 1 1 Chief Scientist 0 1  
## 2 2 Principal Specialist 4 14  
## 3 3 Senior Specialist 46 42  
## 4 4 Specialist 18 10

## Approach to analysis

We are testing whether the frequency distribution of the data is significantly different from an expected frequency distribution. As we do not have equal number of observation in each of the classes (grade) we cannot use a Chi-square test. We will therefore use a G-test of goodness of fit.

The G-test calculates the ratio of observed and expected frequencies and adjusts the ratio for the number of observations. This G value is compared to a chi-square distribution with one fewer degree of freedom than the number of categories. The null hypothesis is that the observed frequencies result from random sampling from a distribution with the given expected frequencies.

## Load packages

We won't need any special R packages for this analysis.

## Compare frequency distributions

### G-test

calculate the expected number of females for each role if the ratios were equal:

grades$femaleExp <- (grades$female + grades$male) / 2  
  
grades

## grade role female male femaleExp  
## 1 1 Chief Scientist 0 1 0.5  
## 2 2 Principal Specialist 4 14 9.0  
## 3 3 Senior Specialist 46 42 44.0  
## 4 4 Specialist 18 10 14.0

#### Calculate G-test for all female staff

#remove first observation (as it is zero and can't be calculated)  
obs <- grades$female[-1]  
expected\_freq <- grades$femaleExp[-1]  
df = length(obs) -1  
lnratio <- log(obs/expected\_freq) \* sum(obs)  
g <- 2\*abs(sum(lnratio))  
g

## [1] 70.06231

# calculate p value   
p.val <- 1-pchisq(g, df)  
p.val

## [1] 6.661338e-16

We tested whether there was a significant difference between the observed distribution of grades among female staff and the expected distribution if there were an equal male and female staff ratio in each grade.

Using a G-test we found a significant difference to the expected distribution (G = 70.0623074, df = 2, p = < 0.001).

#### Calculate G-test for each grade separately

obs <- c(18, 10)  
expected <- c(1, 1)  
expected\_freq <- expected / sum(expected) \* sum(obs)  
lnratio = log(obs/expected\_freq) \* sum(obs)  
g <- 2\*abs(sum(lnratio))  
g

## [1] 4.768837

1-pchisq(q = g, df = df) # degrees of freedom: 2 groups minus 1 = 1

## [1] 0.09214253