Lecture 3

1 Contingency Table Tests for Independence: χ^2

The calculations for the statistical philosophy vs. education example from class are presented here using R.

```
rm(list=ls())
x = matrix(c(90,13,19,12,1,13,78,6,50), nrow=3)
rownames(x) = c("BS","MS","PhD")
colnames(x) = c("Frequentist","Bayesian","Combination")
x
```

```
##     Frequentist Bayesian Combination
## BS           90       12          78
## MS           13        1           6
## PhD          19       13          50
```

1.1 Pearson χ^2 Test

```
pi_hat = x/sum(x)
pi_hat
```

```
##     Frequentist    Bayesian Combination
## BS   0.31914894 0.042553191   0.2765957
## MS   0.04609929 0.003546099   0.0212766
## PhD  0.06737589 0.046099291   0.1773050
```

```
x_marg = apply(x,1,sum)
y_marg = apply(x,2,sum)

mu_hat = (x_marg %*% t(y_marg)) / sum(x)
mu_hat
```

```
##      Frequentist  Bayesian Combination
## [1,]   77.872340 16.595745   85.531915
## [2,]    8.652482  1.843972    9.503546
## [3,]   35.475177  7.560284   38.964539
```

```
X2 = sum((x-mu_hat)^2/mu_hat)
X2
```

```
## [1] 22.37769
```

```
qchisq(.95,(dim(x)[1]-1)*(dim(x)[2]-1))
```

```
## [1] 9.487729
```

We reject the null of independence between statistical philosophy and education because 22.3776853 is greater than 9.487729. You can also use the R function *chisq.test* to obtain the same result.

```
chi2 = chisq.test(x)
chi2
```

```
##
##  Pearson's Chi-squared test
```

```
##
## data:  x
## X-squared = 22.378, df = 4, p-value = 0.0001685
```

1.2 Likelihood Ratio χ^2 Test

```
G2 = 2*(sum(x*log(x/mu_hat)))
G2
```

```
## [1] 23.03619
```

We reject the null of independence between statistical philosophy and education because 23.0361921 is greater than 9.487729. You can also use the R function *GTest* from the *DescTools* pacakge to obtain the same result.

```
library(DescTools)
LRT = GTest(x)
LRT
```

```
##
##  Log likelihood ratio (G-test) test of independence without
##  correction
##
## data:  x
## G = 23.036, X-squared df = 4, p-value = 0.0001245
```

1.3 Nature of Association

The Pearson residuals can be calculated manually or by calling the *residuals* object from the *chisq.test* function.

```
(x - mu_hat) / sqrt(mu_hat)
```

```
##     Frequentist   Bayesian Combination
## BS     1.374312 -1.1281258  -0.8144066
## MS     1.477988 -0.6215137  -1.1364884
## PhD   -2.766100  1.9783686   1.7678919
```

```
chisq.test(x)$residuals
```

```
##     Frequentist   Bayesian Combination
## BS     1.374312 -1.1281258  -0.8144066
## MS     1.477988 -0.6215137  -1.1364884
## PhD   -2.766100  1.9783686   1.7678919
```

The standardized Pearson residuals can be calculated manually or by calling the *stdres* object from the *chisq.test* function.

```
sum(x)*((x - mu_hat) / sqrt(mu_hat))/sqrt(outer(sum(x)-x_marg,sum(x)-y_marg,"*") )
```

```
##     Frequentist   Bayesian Combination
## BS     3.033712 -1.9687318   -1.869216
## MS     2.035680 -0.6767515   -1.627543
## PhD   -4.360557  2.4655927    2.897737
```

```
chisq.test(x)$stdres
```

```
##     Frequentist   Bayesian Combination
## BS     3.033712 -1.9687318   -1.869216
## MS     2.035680 -0.6767515   -1.627543
## PhD   -4.360557  2.4655927    2.897737
```

The greatest differences seem to be the PhD Frequentists (fewer than expected) and the BS Frequentists (more than expected). But all statistical philosophies are quite different than expected for those with a PhD education. We can look at this further via partitioning the G^2 statistic.

```r
GTest(matrix(c(90,13,12,1),nrow=2))
```

```
##
##  Log likelihood ratio (G-test) test of independence without
##  correction
##
## data:  matrix(c(90, 13, 12, 1), nrow = 2)
## G = 0.29419, X-squared df = 1, p-value = 0.5875
```

```r
GTest(matrix(c(102,14,78,6),nrow=2))
```

```
##
##  Log likelihood ratio (G-test) test of independence without
##  correction
##
## data:  matrix(c(102, 14, 78, 6), nrow = 2)
## G = 1.3588, X-squared df = 1, p-value = 0.2437
```

```r
GTest(matrix(c(103,19,13,13),nrow=2))
```

```
##
##  Log likelihood ratio (G-test) test of independence without
##  correction
##
## data:  matrix(c(103, 19, 13, 13), nrow = 2)
## G = 12.953, X-squared df = 1, p-value = 0.0003194
```

```r
GTest(matrix(c(116,32,84,50),nrow=2))
```

```
##
##  Log likelihood ratio (G-test) test of independence without
##  correction
##
## data:  matrix(c(116, 32, 84, 50), nrow = 2)
## G = 8.4303, X-squared df = 1, p-value = 0.00369
```

The null of independence is rejected in subtable 3 and 4 - PhDs are more likely to be Bayesian vs. Frequentist, and PhDs are more likely to be a combination vs. one or the other, respectively.