Lab 8: Chi-square

Frank Woodling

April 27, 2016

1.

H_0 : Boys and girls prefer the same types of sports

H_1 : Boys and girls do not prefer the same types of sports

2.

No, from page 27 of the chapter 5 lecture slides the chi-square distribution may not be valid when some $e_{ij} < 5$. We can still create a permutation distribution of the chi-square statistic though.

3.

The boys row currently adds up to $\frac{3}{5}$, and the girls row adds to $\frac{2}{5}$. Both these would need to equal 1 in order for a traditional test to work properly. We we need at least another 5 students total, 2 boys and 3 girls.

4.

```
source("http://www4.stat.ncsu.edu/~lu/ST505/Rcode/functions-Ch5.R")

sport = c(1,2,2,1,3)
gender = c(1,1,1,2,2)

partytable <- table(gender, sport)
result_obs = chisq.test(partytable)
```

```
## Warning in chisq.test(partytable): Chi-squared approximation may be
## incorrect
```

```
result_obs$statistic
```

```
## X-squared
##  2.916667
```

5.

There are $\binom{5}{3}$ combinations or 10 total combinations.

```
# n     boys       girls
# 1  B1, B2, F1    F2, O1
# 2  B1, B2, F2    F1, O1
# 3  B1, B2, O1    F1, F2
# 4  B1, F1, F2    B2, O2
# 5  B1, F1, O1    B2, F2
# 6  B1, F2, O1    B2, F1
# 7  B1, F1, F2    B2, O1
# 8  B2, F1, O1    B1, F2
# 9  B2, F2, O1    B1, F1
# 10 F1, F2, O1    B1, B2
```

6.

```
a <- matrix(
  c(2, 1,   0, 0,   1, 1,
    2, 1, 0, 0 ,1, 1,
    2, 0 ,1, 0, 2, 0,
    1, 2, 0, 1, 0, 1,
    1, 1, 1, 1, 1, 0,
    1, 1, 1, 1, 1, 0,
    1, 2, 0, 1, 0, 1,
    1, 1, 1, 1, 1, 0,
    1, 1, 1, 1, 1, 0,
    0, 2, 1, 2, 0, 0),
  nrow = 10,
  ncol = 6,
  byrow = T)

dimnames(a) = list(1:10, c("B", "F", "O", "B", "F", "O"))
a
```

```
##    B F O B F O
## 1  2 1 0 0 1 1
## 2  2 1 0 0 1 1
## 3  2 0 1 0 2 0
## 4  1 2 0 1 0 1
## 5  1 1 1 1 1 0
## 6  1 1 1 1 1 0
## 7  1 2 0 1 0 1
## 8  1 1 1 1 1 0
## 9  1 1 1 1 1 0
## 10 0 2 1 2 0 0
```

7.

```
table_chi = rep(0, 10)
for(i in 1:10)
```

```
{
  r1 = a[i, 1:3]
  r2 = a[i, 4:6]
  table = rbind(r1, r2)
  chi_result = chisq.test(table)
  table_chi[i] = chi_result$statistic

}
```

```
## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect

## Warning in chisq.test(table): Chi-squared approximation may be incorrect
```

```
cbind(a, table_chi)
```

```
##    B F O B F O table_chi
## 1  2 1 0 0 1 1 2.9166667
## 2  2 1 0 0 1 1 2.9166667
## 3  2 0 1 0 2 0 5.0000000
## 4  1 2 0 1 0 1 2.9166667
## 5  1 1 1 1 1 0 0.8333333
## 6  1 1 1 1 1 0 0.8333333
## 7  1 2 0 1 0 1 2.9166667
## 8  1 1 1 1 1 0 0.8333333
## 9  1 1 1 1 1 0 0.8333333
## 10 0 2 1 2 0 0 5.0000000
```

8.

```
mean(table_chi >= result_obs$statistic)
```

```
## [1] 0.6
```

This seems like a very high p-value, and may be incorrect. If 0.6 is the p-value then we fail to reject the null hypothesis. We can conclude that boys and girls like the same types of sports.

9.

```
### Make observed contingency table and calculate stat
Row1 = c(22,40,12); Row2 = c(30,17,11)
Table = rbind(Row1,Row2)
teststat.obs = chisq.test(Table)$statistic
teststat.obs
```

```
## X-squared
##  8.744026
```

```
### create the prefernce data and the gender data
preference = c( rep("B",52), rep("F",57), rep("O",23))
gender = c( rep("boy",74), rep("girl",58) )
table(preference); table(gender)
```

```
## preference
##  B  F  O
## 52 57 23
```

```
## gender
##  boy girl
##   74   58
```

```
y = preference; x = gender
teststat = rep(NA, 1000)
for(i in 1:1000) {
### randomly "shuffle" the y data between the x groups
ySHUFFLE = sample(y)
### compute chi-square stat for the shuffled data
TableSHUFFLE = table(x,ySHUFFLE)
teststat[i] = chisq.test(TableSHUFFLE)$statistic
}
### calculate the approximate p-value
mean(teststat >= teststat.obs)
```

```
## [1] 0.009
```

I found a p-value of 0.08 with the full dataset. We can reject the null hypothesis, and conclude that boys and girls prefer different sports.

10.

```
Row1 = c(22,40,12); Row2 = c(30,17,11)
Table = rbind(Row1,Row2)

chisq.test(Table)
```

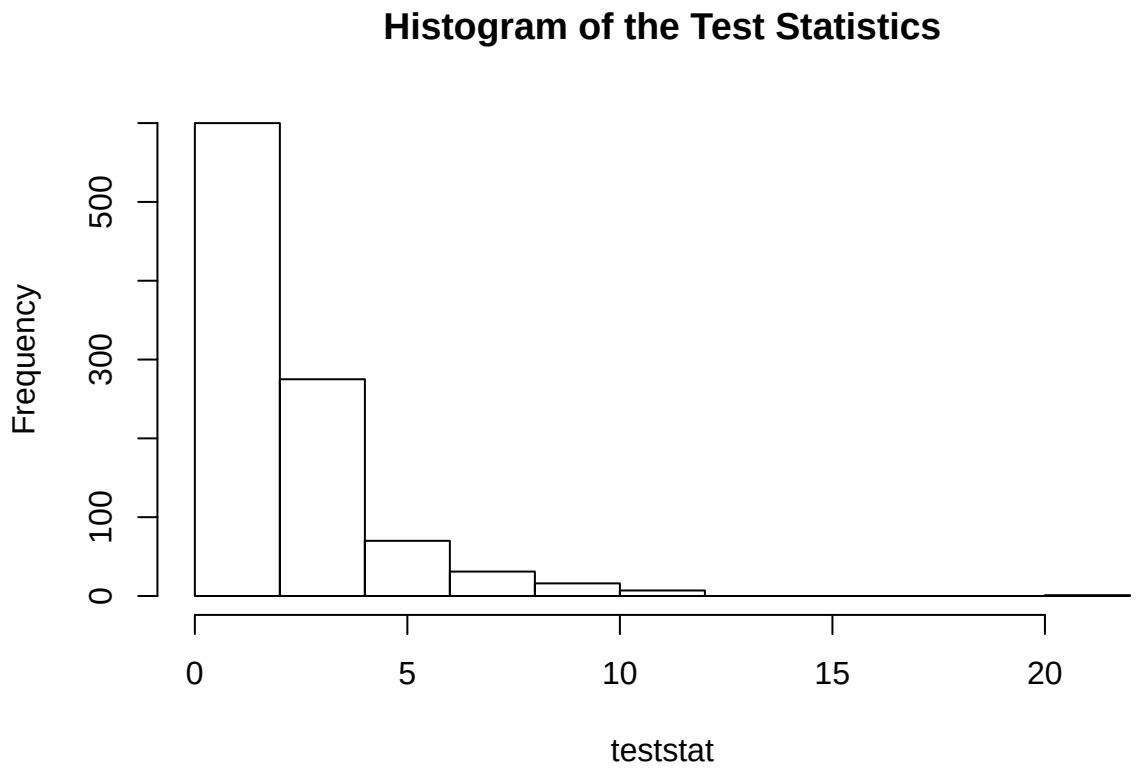
```
##
##  Pearson's Chi-squared test
##
## data:  Table
## X-squared = 8.744, df = 2, p-value = 0.01263
```

Yes, since $e_{ij} > 5$ we could use a traditional test. The test resulted in a p-value of 0.01263 so we can reject the null hypothesis. This value is much smaller than the permutation result.

Summary

1.

The data set contains all of the data for boys and girls preferences of sports between basketball, football, or other sports. For the permutation test we found a test-statistic of 8.74 and a p-value of 0.08, while in the traditional test we found a test-statistic of 8.744 and a p-value of 0.01263. In both cases we can reject the null hypothesis, and conclude boys and girls prefer different sports.



2.

```
#        boys               girls
# 1   B1, B2, F1         F2, O1
# 2   B1, B2, F2           F1, O1
# 3   B1, B2, Soccer       F1, F2
# 4   B1, F1, F2           B2, O2
# 5   B1, F1, Soccer     B2, F2
# 6   B1, F2, Soccer       B2, F1
# 7   B1, F1, F2           B2, O1
# 8   B2, F1, Soccer       B1, F2
# 9   B2, F2, Soccer       B1, F1
# 10    F1, F2, Soccer     B1, B2
# 11    B1, B2, F1         F2, O1
# 12    B1, B2, F2         F1, O1
# 13    B1, B2, Swimming     F1, F2
# 14    B1, F1, F2         B2, O2
# 15    B1, F1, Swimming     B2, F2
# 16    B1, F2, Swimming     B2, F1
# 17    B1, F1, F2         B2, O1
# 18    B2, F1, Swimming     B1, F2
# 19    B2, F2, Swimming     B1, F1
# 20    F1, F2, Swimming     B1, B2
# 21    B1, B2, F1         F2, O1
# 22    B1, B2, F2         F1, O1
# 23    B1, B2, Track        F1, F2
# 24    B1, F1, F2         B2, O2
# 25    B1, F1, Track        B2, F2
# 26    B1, F2, Track        B2, F1
# 27    B1, F1, F2         B2, O1
# 28    B2, F1, Track        B1, F2
# 29    B2, F2, Track        B1, F1
# 30    F1, F2, Track        B1, B2
```

I believe this would create a total of 30 combinations since there is only 1 other in each permutation between the boys and girls. Each other could be split into the 3 new sports.