STAT 3480 Lab8

Shaoran Sun

April 26, 2016

Part 1

H_0: There is no association between the row and the column factors. (Gender and sport preference are independent)

H_a: There is association between the row and the column factors. (Gender and sport preference are dependent)

Part 2

No. Because the expected counts of every cell in the table are less than 5.

	basketball	football	other	total
boys	1	2	0	3
girls	1	0	1	2
total	2	2	1	5

Table of Expected Counts

	basketball	football	other	total
boys	6/5	6/5	3/5	3
girls	4/5	4/5	2/5	2
total	2	2	1	5

Part 3

We need at least 63 students.

Since the least expected count in the table is "girls other", we need the expected count of this cell greater than 5, so number others should be greater than $5 \div \frac{2}{5}$, which is 12.5. This is 1/5 of the total students. Hence, we have a total number of students of $12.5/(1/5) = 62.5$

Part 4

$\chi^2_{obs} = 2.9167$

Part 5

There are 10 rows in the table.

boys	girls
B1 B2 F1	F2 O1
B1 B2 F2	F1 O1
B1 B2 O1	F1 F2
B1 F1 F2	B2 O1
B1 F1 O1	B2 F2
B1 F2 O1	B2 F1
B2 F1 F2	B1 O1
B2 F1 O1	B1 F2
B2 F2 O1	B1 F1
F1 F2 O1	B1 B2

Part 6

boys B F O	girls B F O
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

Part 7

boys B F O	girls B F O	test statistic χ^2
2 1 0	0 1 1	2.9167
2 1 0	0 1 1	2.9167
2 0 1	0 2 0	5
1 2 0	1 0 1	2.9167
1 1 1	1 1 0	0.83333
1 1 1	1 1 0	0.83333
1 2 0	1 0 1	2.9167
1 1 1	1 1 0	0.83333
1 1 1	1 1 0	0.83333
0 2 1	2 0 0	5

Part 8

χ^2_{obs} is 2.9167. Out of the ten permutations, 6 are greater or equal to 2.9167. Hence, the p-value is $6/10 = 0.6$, which is greater than 0.05. Therefore we fail to reject the null hypothesis and conclude that there is not enough evidence to show gender and sport preference are dependent.

Part 9

χ^2_{obs} is 8.744026. Out of the 1000 random permutations, the p-value is 0.013, which is less than 0.05. Therefore we reject the null hypothesis and conclude that gender and sport preference are dependent.

Part 10

Yes. The least expected count is 10.11, which is greater than 5. So we can apply traditional method. We have the same test statistic of 8.744, and similar p-value of 0.013, which is still less than 0.05. Therefore we still reject the null hypothesis and conclude that gender and sport preference are dependent.

	basketball	football	other	total
boys	22	40	12	74
girls	30	17	11	58
total	52	57	23	132

Table of Expected Counts

	swimming	soccer	track	total
boys	4	4	4	12
girls	4	4	3	11
total	8	8	7	23 (grand total = 132)

Appendix

4

```
A = matrix( c(1,2,0,1,0,1), # the data elements
            nrow=2,              # number of rows
            ncol=3,              # number of columns
            byrow = TRUE)        # fill matrix by rows

dimnames(A) = list( c("boys", "girls"),         # row names
                    c("basketball", "football", "other")) # column names
chisq.test(A)
```

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

```
##
##  Pearson's Chi-squared test
##
## data:  A
## X-squared = 2.9167, df = 2, p-value = 0.2326
```

7

```r
chisq.test(rbind(c(2 ,1 ,0),c( 0, 1, 1)))
```

```
## Warning in chisq.test(rbind(c(2, 1, 0), c(0, 1, 1))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(2, 1, 0), c(0, 1, 1))
## X-squared = 2.9167, df = 2, p-value = 0.2326
```

```r
chisq.test(rbind(c(2, 1, 0),c( 0, 1, 1)))
```

```
## Warning in chisq.test(rbind(c(2, 1, 0), c(0, 1, 1))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(2, 1, 0), c(0, 1, 1))
## X-squared = 2.9167, df = 2, p-value = 0.2326
```

```r
chisq.test(rbind(c(2 ,0 ,1),c( 0 ,2 ,0)))
```

```
## Warning in chisq.test(rbind(c(2, 0, 1), c(0, 2, 0))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(2, 0, 1), c(0, 2, 0))
## X-squared = 5, df = 2, p-value = 0.08208
```

```r
chisq.test(rbind(c(1, 2 ,0),c( 1 ,0 ,1)))
```

```
## Warning in chisq.test(rbind(c(1, 2, 0), c(1, 0, 1))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(1, 2, 0), c(1, 0, 1))
## X-squared = 2.9167, df = 2, p-value = 0.2326
```

```r
chisq.test(rbind(c(1, 1 ,1),c(  1 ,1 ,0)))
```

```
## Warning in chisq.test(rbind(c(1, 1, 1), c(1, 1, 0))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(1, 1, 1), c(1, 1, 0))
## X-squared = 0.83333, df = 2, p-value = 0.6592
```

```
chisq.test(rbind(c(1, 1 ,1),c(  1 ,1 ,0)))
```

```
## Warning in chisq.test(rbind(c(1, 1, 1), c(1, 1, 0))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(1, 1, 1), c(1, 1, 0))
## X-squared = 0.83333, df = 2, p-value = 0.6592
```

```
chisq.test(rbind(c(1, 2 ,0),c(  1 ,0 ,1)))
```

```
## Warning in chisq.test(rbind(c(1, 2, 0), c(1, 0, 1))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(1, 2, 0), c(1, 0, 1))
## X-squared = 2.9167, df = 2, p-value = 0.2326
```

```
chisq.test(rbind(c(1, 1 ,1),c(  1 ,1 ,0)))
```

```
## Warning in chisq.test(rbind(c(1, 1, 1), c(1, 1, 0))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(1, 1, 1), c(1, 1, 0))
## X-squared = 0.83333, df = 2, p-value = 0.6592
```

```
chisq.test(rbind(c(1, 1 ,1),c(  1 ,1 ,0)))
```

```
## Warning in chisq.test(rbind(c(1, 1, 1), c(1, 1, 0))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(1, 1, 1), c(1, 1, 0))
## X-squared = 0.83333, df = 2, p-value = 0.6592
```

```r
chisq.test(rbind(c(0, 2 ,1),c(  2 ,0, 0)))
```

```
## Warning in chisq.test(rbind(c(0, 2, 1), c(2, 0, 0))): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(c(0, 2, 1), c(2, 0, 0))
## X-squared = 5, df = 2, p-value = 0.08208
```

9

```r
### 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
```

```r
### 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
```

```r
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
sum(teststat >= teststat.obs)/1000
```

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

```r
sum(teststat >= teststat.obs)
```

```
## [1] 9
```

10

```r
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
```