

Chi-Square Lab Summary

Shaoran Sun

April 26, 2016

Part 1

The null hypothesis is that there is no association between the row and the column factors, in another words, gender and sport preference are independent. The alternate hypothesis is that there is association between the row and the column factors, in another words, gender and sport preference are dependent.

The observed teststat is 8.74. The p -value for traditional test is 0.01263, and for permutation test is 0.013.

Please see the [last page](#) for the histogram.

Sicne the p -value is less than 0.05, we reject the null hypothesis and conclude that gender and sport preference are dependent.

Part 2

If “other” category is broken into 3 subcategories, the expected counts of any cell for the 3 new subcategories would be less than 5. Let’s take the max gender total, boys total, which is 12, and multiply $(4 + 4 = 8)$, which is the average sum of each of the three subcategories: $12 \times 8 = 96$. $96/132 < 5$. If the average is less than 5, then, any other subcategories would not have expected count that is greater than 5. Because if one is greater than the average sum, 8, then there must be another cell that is less than 8. Hence, I would not feel comfortable performing a traditional chi-square test if “other” category is broken into 3 subcategories.

Histogram

