# Stat 4201 Homework 3

Mengqi Zong < $mz2326@columbia.edu$ >

February 10, 2012

## Problem 1

I use Fisher's Exact Test to do the data analysis. Here is the result from R:

```
Fisher's Exact Test for Count Data

data:  matrix(c(149, 48, 129, 68), 2, 2)
p-value = 0.03547
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.032059 2.602066
sample estimates:
odds ratio
  1.634216
```

The proportion of incorrect recollections among the left-handed (75.6%) was higher than the correspond proportion among the right-handed. And the data are not consistent with the hypothesis of equal proportions of incorrect recollections in the populations of left-handed and righ-handed people ($two - sidedp - value = .035$). That is, the data indicate that left- or right-handedness is associated with correct reollection of the orientation. And the 95% confidence interval for the odds ratio is $[1.032, 2.602]$. This means that left-handed people is 1.032 times to 2.602 times higher than right-handed people to recollect the comet orientation incorrectly.

# Problem 2

The Simpson's paradox is caused by the different proportion of math and physics majors. In this example, the paradox is caused by Eagle College's high proportion of Mathematics majors, which has a higher pass rate than that of both colleges' Physics majors.

Among math students, 67% passed at Crane, compared to 56% at Eagle. Among physics students, 36% passed at Crane, compared to 25% at Eagle. Among all students (ignoring department), 50% passed at Crane compared at 53% at Eagle. Intuitively, both college's average pass rate are affected by the poor pass rate of the physics majors severely. However, due to Eagle college's low proportion of physics majors, its average pass rate doesn't "drop" from the original math major pass rate much (from 56% to 53%). On the contrary, Crane's does (from 67% to 50%). As a result, Simpson's paradox arises.

However, Crane still won, though its average pass rate is less than that of Eagle's. Because Crane won both match-ups involving students with supposedly comparable training. And what average pass rate did, trying to compare pass between different majors which are not comparable, is not reasonable.

* To answer this question, I referenced the solution at page 578 of Ramsey and Schafer $2^{nd}$ edition.

# Problem 3

I use the Mantel-Haenszel test to do the data analysis. Here is the result from R:

```
Mantel-Haenszel chi-squared test with continuity correction

data:  data.p3
Mantel-Haenszel X-squared = 17.4618, df = 1, p-value = 2.931e-05
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 0.1296716 0.4806826
```

```
sample estimates:
common odds ratio
        0.2496615
```

From the analysis, we can see that the odds of a tire-related fatal accident depend on whether the sports utility vehicle is a Ford ($p - value = 2.931e - 05$). And the 95% confidence interval for odds ratio between Ford and other is $[0.1297, 0.4807]$. Also, the common odds ratio is 0.2497. This means that Ford's sports utility vehicles will cause 4 ($1/0.2497 \approx 4$) times tire accidents than other brands' vehicles.

# Appendices

The code is listed below:

```
# problem 1

fisher.p1 <- fisher.test(matrix(c(149,48,129,68),2,2))
print(fisher.p1)

# problem 3

data.p3 <-
array(c(171, 465, 0, 2,
        243, 753, 0, 3,
         98, 325, 3, 3,
        108, 273, 21, 8),
      dim = c(2, 2, 4),
      dimnames = list(
          Make = c("Ford", "Other"),
          Cause = c("Other", "Tire"),
          Speed = c("0-40", "41-55", "56-65", ">66")))

mantelhaen.p3<- mantelhaen.test(data.p3)
print(mantelhaen.p3)
```