Quiz #2

Dustin Leatherman

May 5, 2019

1

The following table is from a report on the relationship between aspirin use and myocardial infarction (heart attacks) by the Physicians' Health Study Research Group at Harvard Medical School.

Myocardial Infarction

Group	Yes	No	Total
Placebo	189	10845	11034
Aspirin	104	10933	11037

The Physicians' Health Study was a five-year randomized study testing whether regular intake of aspirin reduces mortality from cardiovascular disease. Every other day, the male physicians participating in the study took either one aspirin tablet or a placebo. They study was "blind" – the physicians in the study did not know which type of pill they were taking.

a

Compute and interpret a 95% confidence interval for the difference in proportion of myocardial infarction between the two treatments.

$$\hat{\pi}_2 - \hat{\pi}_1 \pm z_{0.975}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_2)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_1)}{n_2}}$$

$$\frac{104}{11037} - \frac{189}{11034} \pm 1.96\sqrt{\frac{0.0169(1-0.0095)}{11034} + \frac{0.0095(1-0.0169)}{11037}}$$

UCL = -0.0047 LCL = -0.0107

With 95% confidence, Patients who took Aspirin were between 0.47% and 1.07% less likely to have a heart attack than Patients who took a Placebo.

b

Compute and interpret a 95% confidence interval for the odds ratio comparing the risk of myocardial infarction between the placebo and aspirin treatment groups.

```
theta <- (189 * 10933) / (104 * 10845)
n1 <- 11034
n2 <- 11037
pi1 <- 189 / 11034
```

```r
pi2 <- 104 / 11037
ucl <- log(theta) + qnorm(0.975) * sqrt((1 / (n1 * pi1 * (1 - pi1))) + (1 / (n2 * pi2 * (1 - pi2))))
lcl <- log(theta) - qnorm(0.975) * sqrt((1 / (n1 * pi1 * (1 - pi1))) + (1 / (n2 * pi2 * (1 - pi2))))

c(exp(lcl), exp(ucl))
```

```
## [1] 1.440042 2.330780
```

With 95% confidence, Patients who took took the Placebo were between 1.44 and 2.3308 times more likely to receive a heart attack than those who took Aspirin.

2

Is there a gender gap in political affiliation? The following table uses data from the 2000 General Social Survey. It cross classifies gender and political party identification more strongly with the Democratic or Republican party or as Independents.

**

**

Perform a test on whether gender and party identification are independent factors. State the null and alternative hypotheses, the test statistic, the p-value, and the conclusion. Do not forget to check the assumptions of the test.

The appropriate test for checking for independence for an R x C table between Gender and Party Identification is the Chi-Squared Mantel-Haentzsel Test.

```r
# short answer
parties <- data.frame(Democrat=c(762,484), Independent=c(327,239), Republican=c(468,477), row.names = c
chisq.test(parties)$expected %>% kable
```

	Democrat	Independent	Republican
Females	703.6714	319.6453	533.6834
Males	542.3286	246.3547	411.3166

```r
chisq.test(parties)
```

```
##
##  Pearson's Chi-squared test
##
## data:  parties
## X-squared = 30.07, df = 2, p-value = 2.954e-07
```

```r
# long answer
r1c1 <- (762 - ((1246 * 1557) / 2757))^2 / ((1246 * 1557) / 2757)
r1c2 <- (327 - ((566 * 1557) / 2757))^2 / ((566 * 1557) / 2757)
r1c3 <- (468 - ((945 * 1557) / 2757))^2 / ((945 * 1557) / 2757)

r2c1 <- (484 - ((1246 * 1200) / 2757))^2 / ((1246 * 1200) / 2757)
```

```r
r2c2 <- (239 - ((566 * 1200) / 2757))^2 / ((566 * 1200) / 2757)
r2c3 <- (477 - ((945 * 1200) / 2757))^2 / ((945 * 1200) / 2757)

test.chi <- sum(r1c1, r1c2, r1c3, r2c1, r2c2, r2c3)

1 - pchisq(test.chi, (3 - 1) * (2 - 1))
```

```
## [1] 2.953589e-07
```

The expected count size is greater than 5 for all cells so a chi-square test can be run.

H_0: Gender and Party Identification are Independent H_A: There is a dependence between Gender and Party Identification

$\chi^2 = 30.07$

There is convincing evidence that Gender and Political Affiliation are **dependent** (p-value = 2.954e-07).

3

When drinking tea, a colleague of Ronald Fisher's at Rothamsted Experiment Station near London claimed she could distinguish whether milk or tea was added to the cup first. To test her claim, Fisher designed an experiment in which she tasted eight cups of tea. Four cups had milk added first, and the other four had tea added first. She was told there were four cups of each type and she should try to select the four that had milk added first. The cups were presented to her in random order. A potential result of the experiment is presented in the table below.

Perform a test for a positive association between the true order of pouring and her guess. What type of test will you use?

The Fisher Test should be used to test whether the true odds between Poured First and Guess Poured First are equal to 1.

```r
tea <- data.frame(Milk=c(3,1), Tea=c(1,3), row.names = c("Milk","Tea"))
fisher.test(tea)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  tea
## p-value = 0.4857
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##    0.2117329 621.9337505
## sample estimates:
## odds ratio
##   6.408309
```

There is no convincing evidence that the drink Poured First changed the odds for the First Guess (p-value = 0.4857).