STAT 205 Homework 10 Solution

Instructor: Shiwen Shen

Total Points: 100

1. Using in data in Exercise 10.2.5 in page 392 to conduct a chi-square test **without direction** using R. Clearly state null and alternative hypotheses, make decision with 95% confidence level, and write proper interpretation. (10)

   The null and alternative hypotheses are $H_0: p_1 = p_2$ and $H_A: p_1 \neq p_2$, where p_1 is the probability to have wilt disease in Mites group, and p_2 is the probability to have wilt disease in No milts group. We use the following R code to find p-value:

```
> plant <- matrix(c(11,15,17,4), nrow=2)
> colnames(plant) <- c("Mites", "No mites")
> rownames(plant) <- c("Wilt disease", "No wilt disease")
> plant
                Mites No mites
Wilt disease       11       17
No wilt disease    15        4
> chisq.test(plant, correct=FALSE)

        Pearson's Chi-squared test

data:  plant
X-squared = 7.2037, df = 1, p-value = 0.007275

> chisq.test(plant)

        Pearson's Chi-squared test with Yates' continuity correction

data:  plant
X-squared = 5.6885, df = 1, p-value = 0.01708
```

   With(without) using the Yates' continuity correction, the p-value is 0.007(0.017). Both of them are smaller than 5% significance level. (95% confidence level leads to 100%-95%=5% significance level.) Therefore, we reject the null hypothesis and say the spider mites would **change** the probability to have wilt disease for potted cotton.

2. Using the same data in problem 1. (1) Conduct a Fisher's exact test on whether mites induces resistance to wilt disease using R. Note that this is a test **with direction**. Clearly state null and alternative hypotheses, make decision with 95% confidence level, and write proper interpretation. (2) Compare to problem 1, which test is more powerful in rejecting the null hypothesis? There are two reasons to explain one of them is more powerful. Explain here. (3) List **four** "more extreme" cases if we want to calculate the p-value by hand. (You need to show me 4 tables.) (30)

   (1) The null and alternative hypotheses are $H_0: p_1 = p_2$ and $H_A: p_1 < p_2$, where p_1 is the probability to have wilt disease in Mites group, and p_2 is the probability to have wilt disease in No milts group. The p-value for Fishers' exact test can be achieved by the following R code:

```
> fisher.test(plant, alternative="less")

        Fisher's Exact Test for Count Data

data:  plant
p-value = 0.007743
```

   Since p-value is less than 5% level of significance, we reject the null hypothesis and claim spider mites would **decrease** the probability to have wilt disease for potted cotton.

   (2) Second test is more power because it has a direction, and a test with direction is always more powerful in rejecting the null hypothesis. Besides, the chi-square test is an approximate test. Instead Fishers' exact test calculates the exact value of p-value by enumerating all possible "more extreme" cases, which always more powerful.

   (3) The four tables are:

10	18
16	3

9	19
17	2

8	20
18	1

7	21
19	0

3. Using the data in Exercise 10.5.8 in page 412. (1) Conduct a chi-square test in R. Clearly state null and alternative hypotheses, make decision with 95% confidence level, and write proper interpretation. (2) Calculate the degree-of-freedom by hand. (20)

(1) The null hypothesis is

$$H_0: p_1 = p_2 = p_3 = p_4$$

Where p_1 is the probability of receiving Angioplasty in Cardiac death group, p_2 is the one in Other death group, p_3 is the one in Unknown case of death group, and p_4 is the one in Alive group. The alternative is H_A: at least one of p_1, p_2, p_3 and p_4 is different.

```
> patient <- matrix(c(23,25,45,51,17,19,1064,1043), nrow=2)
> colnames(patient) <- c("Cardiac", "Other", "Unknown", "Alive")
> rownames(patient) <- c("Angioplasty", "Medical therapy")
> patient
                Cardiac Other Unknown Alive
Angioplasty          23    45      17  1064
Medical therapy      25    51      19  1043
> chisq.test(patient)

        Pearson's Chi-squared test

data:  patient
X-squared = 0.7259, df = 3, p-value = 0.8671
```

With the above R output, we find the p-value is 0.8671 > 0.05. Therefore, we accept the null hypothesis and claim that there is no association between treatment group and whether receiving additional Angioplasty.

(2) The degree-of-freedom is (2-1)(4-1)=(1)(3)=3.

4. Exercise 10.7.3 in page 418. Use R. (10)

Based on the following R output, the 95% confidence interval for Pr{preterm | bed rest} – Pr{preterm | control} is (0.003, 0.233). Since "0" is not covered by the confidence interval and the lower bound is positive, we can conclude with 95% confidence that bed rest is not beneficial to reduce the risk of premature delivery with twin pregnancies, instead it increases the risk!

```
> total <- c(105, 107)
> preterm <- c(32, 20)
> prop.test(preterm, total, correct=FALSE)

        2-sample test for equality of proportions without continuity
        correction

data:  preterm out of total
X-squared = 3.9757, df = 1, p-value = 0.04616
alternative hypothesis: two.sided
95 percent confidence interval:
 0.002919857 0.232772177
```

5. Exercise 10.9.3 in page 429. (5)

$$Relative\ Risk(RR) = \frac{3995/46941}{221/5228} = 2.013$$

6. Exercise 10.9.4 in page 429. Use R in (b). (10)

(a) Using the formula in Notes 19 page 9, the odds ratio is

$$Odds\ Ratio(OR) = \frac{3995 \times 5007}{221 \times 42946} = 2.108$$

(b) The 95% confidence interval for the population value of the odds ratio is (1.834, 2.432) based on the R output.

```
> dog <- matrix(c(3995,42946,221,5007), nrow=2)
> rownames(dog) <- c("Yes", "No")
> colnames(dog) <- c("Golden retriever", "Border collie")
> dog
    Golden retriever Border collie
Yes             3995           221
No             42946          5007
> fisher.test(dog, conf.int=TRUE)

        Fisher's Exact Test for Count Data

data:  dog
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.834033 2.432048
```

(c) We are 95% confident that Golden retriever has the higher odds of hip dysplasia by 1.834 to 2.432 times, relative to Border collie.

7. Switch two columns in the table in Exercise 10.9.3. (1) Calculate the odds ratio and make a proper interpretation of the odds ratio. (2) Compare the odds ratio you calculated in problem 6, what is the relationship between two odds ratios? (15)

(1) After switching, the table becomes

	Border collie	Golden retriever
Yes	221	3995
No	5007	42946
Total	5228	46941

Using the formula in Notes 19 page 9, the odds ratio is

$$Odds\ Ratio(OR) = \frac{221 \times 42946}{3995 \times 5007} = 0.474$$

Interpretation: the odds of having hip dysplasia for Border collie are 0.474 times of the one of Golden retriever.

(2) The relation between 0.474 and 2.108 is $0.474 = \frac{1}{2.108}$.