STAT 545 Homework 5

Quan Zhou*

October 4, 2015

1. **Problem 3.1** (2 points)
   I think usually given such a table, people would like to know whether fastening the seat belt could decrease the risk of fatal injury in accidents. We will use the asymptotic results from Section 3.2.3 of the textbook to answer this question.

   (a) Odds ratio (obviously we need no continuity correction here):

$$\begin{aligned}
\hat{\theta} &= \frac{1601 \times 412368}{510 \times 162527} = 7.96 \\
\log \hat{\theta} &= 2.075 \\
ASE(\log \hat{\theta}) &= (\frac{1}{1601} + \frac{1}{510} + \frac{1}{162527} + \frac{1}{412368})^{1/2} = 0.051 \\
95\% \text{ CI } (\theta) &= (e^{1.976}, e^{2.175}) = (7.21, 8.80)
\end{aligned}$$

   From the table, we can estimate that the odds of fatal injury when the seat belt is not used is 7.96 times that when the seat belt is used. Asymptotically $\log \hat{\theta}$ follows a normal distribution with standard deviation 0.051. Hence we have 95% confidence that the true value of odds ratio is in $(7.21, 8.80)$.

   (b) Difference of proportions:

$$\begin{aligned}
p_{1|1} - p_{1|2} &= \frac{1601}{1601 + 162527} - \frac{510}{510 + 412368} = 0.00852 \\
ASE(p_{1|1} - p_{1|2}) &= [\frac{1601 \times 162527/(1601 + 162527)^2}{1601 + 162527} + \frac{510 \times 512368/(510 + 412368)^2}{510 + 412368}]^{1/2} \\
&= 0.000249 \\
95\% \text{ CI } (p_{1|1} - p_{1|2}) &= (0.00852 - 1.96 \times 0.000249, 0.00852 + 1.96 \times 0.000249) \\
&= (0.00803, 0.00901)
\end{aligned}$$

   From the table, we can estimate that the difference between the chance of fatal injury when the seat belt is not used and that when the seat belt is used is 0.00852. We are 95% confident that the true value of this difference is in $(0.00803, 0.00901)$.

*an adapted version from Quan's original homework.

(c) Relative risk:

$$\begin{aligned}
\hat{r} &= \frac{1601/(1601+162527)}{510/(510+412368)} = 7.90 \\
\log\hat{r} &= 2.066 \\
ASE(\log\hat{r}) &= (\frac{1}{1601} + \frac{1}{510} - \frac{1}{1601+162527} - \frac{1}{510+412368})^{1/2} = 0.0508 \\
95\% \text{ CI } (r) &= (e^{1.967}, e^{2.166}) = (7.15, 8.72)
\end{aligned}$$

From the table, we can estimate that the risk of fatal injury when the seat belt is not used is 7.90 times that when the seat belt is used. Asymptotically logarithm of the relative risk $\log\hat{r}$ follows a normal distribution with standard deviation 0.0508. Hence we have 95% confidence that the true value of relative risk is in $(7.21, 8.80)$.

2. **Problem 3.4 (a)** (2 points)

(a) Pearson's chi-squared test. First we compute the expected values assuming independence. They are shown in the following table.

Race	Democrat	Independent	Republican
Black	58.44	15.80	54.76
White	385.56	105.20	361.24

Hence,

$$X^2 = \sum\sum \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = 79.431$$

Since $d.f. = (3-1)\times(2-1) = 2$ here, $X^2 \xrightarrow{D} \chi^2_2$. So P-value is

$$P = 5.65 \times 10^{-18}$$

(b) Likelihood ratio test.

$$G^2 = 2\sum\sum n_{ij}\log(n_{ij}/\hat{n}_{ij}) = 90.33$$

Still we have $d.f. = 2$. So,

$$P = 2.43 \times 10^{-20}$$

We see that the two tests have the same asymptotic distribution. Since both methods are asymptotic, it is no surprise that the P-values of two tests differ, especially considering that the sample size we have is not very large. However, both tests give extremely small P-values that suggest the null hypothesis should be rejected. So our conclusion is party identification is not independent of race.

3. **Problem 3.4 (c)** (2 points)

The test we did in the last problem could be partitioned into two independent parts.

Sub-table 1

Race	Democrat	Independent
Black	103(92.9)	15(25.1)
White	341(351.1)	105(94.5)

Sub-table 2

Race	Democrat + Independent	Republican
Black	118(74.2)	11(54.8)
White	446(489.8)	405(361.2)

So we can do likelihood ratio test for each of them.

$$G_1^2 = 2\sum\sum n_{ij}\log(n_{ij}/\hat{n}_{ij}) = 7.16$$
$$G_2^2 = 2\sum\sum n_{ij}\log(n_{ij}/\hat{n}_{ij}) = 83.17$$

Each of them asymptotically follows a χ_1^2 distribution. So

$$P_1 = 0.00747$$
$$P_2 = 7.53 \times 10^{-20}$$

We notice that the sum of two statistics is exactly equal to the G^2 in the last problem, which in some sense proves the independence of the two sub-tests and reveals the value of partitioning. Now we can clearly see that the difference of party identification of two races is mostly reflected by their attitudes towards Republican. Although there is also arguably strong evidence for dependence in the first sub-test, the main distinction is that White supports Republican much more than Black.

4. **Problem 3.14** (2 points)
We first tabulate the information.

Group	Normalized	Non-normalized
Treatment	7	8
Control	0	15

Of course we can do a Pearson's chi-squared test with continuity correction. However, given the limited sample size, we would prefer a Fisher's exact test. The probability of observing this table is

$$Pr = \frac{\binom{7}{7}\binom{23}{8}}{\binom{30}{15}} = 0.00316$$

Clearly this is already the most extreme case given that the row sums and the column sums are fixed. Hence the P-value for a one-sided Fisher's exact test is

$$P = 0.00316 < 0.05$$

Hence we reject the null hypothesis. We claim there is strong evidence showing that the treatment is truly effective.