2019R1 Discrete Data Analysis (STAT5107) Assignment 2

Yiu Chung WONG 1155017920

```r
set.seed(5107);
```

1a.

- relative risk

1bi.

```r
(1-.45) / 1
```

```
## [1] 0.55
```

1bii.

```r
1 / (1-.45)
```

```
## [1] 1.818182
```

2.

$$
\begin{aligned}
P(\text{disease} \mid +) &= \frac{P(+ \mid \text{disease})P(\text{disease})}{P(+)} \\
&= \frac{P(+ \mid \text{disease})P(\text{disease})}{P(+ \mid \text{disease})P(\text{disease}) + P(+ \mid \text{healthy})P(\text{healthy})} \\
&= \pi_1 p/[\pi_1 p + \pi_2(1-p)]
\end{aligned}
$$

3.

- Injury

```r
table <- matrix(data = c(1601, 510, 162527, 412368), 2, 2,
                dimnames = list(c("None", "Seat belt"), c("Fatal", "Nonfatal")));
pi_table <- prop.table(as.matrix(table), margin = 1)
difference_of_proportions <- pi_table[1,1] - pi_table[2,1];
relative_risk <- pi_table[1,1] / pi_table[2,1];
odds_ratio <- (pi_table[1,1] * pi_table[2,2]) / (pi_table[1,2] * pi_table[2,1]);
```

- difference of proportions = 0.0085194
- relative risk = 7.8969651
- odds ratio = 7.9649049
- Because both π_{11} and π_{12} are close to zero

4a.

```
table <- matrix(data = c(28, 18, 656, 658), 2, 2,
                dimnames = list(c("placebo", "Aspirin"), c("Yes", "No")));
pi_table <- prop.table(as.matrix(table), margin = 1)

sample_odds_ratio <- (table[1,1] * table[2,2]) / (table[2,1] * table[1,2]);

se_log_theta <- sqrt(sum(1/table));

CI_log_theta <- qnorm(c(.025, .975), log(sample_odds_ratio), se_log_theta);
CI_theta <- exp(CI_log_theta);
```

- Sample odds ratio is 1.5602981
- standard error of $log\hat{\theta}$ is 0.3071058
- 95% confidence interval for $log\theta$ lies between -0.1570395 and 1.0467933
- 95% confidence interval for θ lies between 0.8546703 and 2.848502
- The 95% confidence interval includes the value 1. Hence it is possible for the two to be equal.

4b.

```
exp_diff <- pi_table[1,1] - pi_table[2,1];
var <- pi_table[1,1] * pi_table[1,2] / sum(table[1,]) +
  pi_table[2,1] * pi_table[2,2] / sum(table[2,]);
sd <- sqrt(var);
wald_CI <- qnorm(c(.025, .975), exp_diff, sd);
```

- 95% confidence interval for difference of proportions lies between -0.0022472 and 0.016953

4c.

```
sample_relative <- pi_table[1,1] / pi_table[2,1];
exp_log_relative <- log(sample_relative);
var <- ((1 - pi_table[1,1]) / ((sum(table[1,])) * pi_table[1,1])) +
  ((1 - pi_table[2,1]) / ((sum(table[2,])) * pi_table[2,1]))
sd <- sqrt(var);
wald_CI_log <- qnorm(c(.025, .975), exp_log_relative, sd);
wald_CI <- exp(wald_CI_log);
```

- Sample relative risk = 1.5373619
- 95% confidence interval for the log relative risk lies between -0.1524358 and 1.0125716
- 95% confidence interval for the relative risk lies between 0.858614 and 2.7526707

5.

H_0: independence for all combination H_1: at least one combination not independent

```
table <- matrix(data = c(178, 570, 138, 138, 648, 252, 108, 442, 252), 3, 3,
                dimnames = list(c("Less than high school", "High school or junior college",
                                  "Bachelor or graduate"),
```

```r
                                  c("Fundamentalist", "Moderate", "Liberal")));

chisq.test(table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 69.157, df = 4, p-value = 3.42e-14
```

- p-value is zero; there are enough evidence to reject H_0; there is an association.