STAT 477/577 - Technology Guide

Module 2 - Section 3

Relative Risks and Odds Ratios

Below is an explanation of the R commands and functions needed to calculate the relative risk and odds ratio with confidence intervals.

- **Relative Risk**

  There is no built-in function in R for calculating the relative risk and obtaining its confidence interval. I wrote a function called `rr.ci` that will calculate the sample relative risk and the confidence interval for the population relative risk. The general command is:

  `rr.ci(y, n, conf.level)`

  where `y` is the vector of the number of successes in the two groups and `n` is the vector of sample sizes in the two groups. The level of confidence, `conf.level`, should be expressed as a proportion. The function's output is self-explanatory.

  For the prostate cancer example from the lecture notes, we first need to read in the data from the file **cancersurgery.csv**.

```
surgery.data<- read.csv(file.choose(), header = T)
```

  Then we need set the category `No` as the baseline category for both variables.

```
surgery.data$Surgery<- factor(surgery.data$Surgery,
                              levels = c("Yes", "No"))
surgery.data$Died<- factor(surgery.data$Died,
                           levels = c("Yes", "No"))
```

  We will then use R to calculate the contingency table and the number of patients in each group.

```
surgery.table<- table(surgery.data$Surgery, surgery.data$Died)
groups<- margin.table(surgery.table, 1)
```

  Finally the calculation of the sample relative risk estimate and confidence interval for the population relative risk is:

```
rr.ci(surgery.table[,1], groups, conf.level = 0.95)

## Estimated Relative Risk =  0.5176164
## Confidence Interval for Population Relative Risk =  0.2884473 0.9288588
```

- **Odds Ratio**

  There is no built-in function in R for calculating the odds ratio and obtaining its confidence interval in R. I wrote a function called `or.ci` that will calculate the sample odds ratio and the confidence interval for the population odds risk. The general command is:

  `or.ci(y, n, conf.level)`

  where `y` is the vector of the number of successes in the two groups and `n` is the vector of sample sizes in the two groups. The level of confidence, `conf.level`, should be expressed as a proportion. The function's output is self-explanatory.

  For the doctor's survey data from the lecture notes, we first need to read in the data from the file **doctorsurvey.csv**.

```
survey.data<- read.csv(file.choose(), header = T)
```

  Then we need set the category `No` as the baseline category for both variables.

```
survey.data$Receive.Letter<- factor(survey.data$Receive.Letter,
                          levels = c("Yes", "No"))
survey.data$Return.Survey<- factor(survey.data$Return.Survey,
                          levels = c("Yes", "No"))
```

  We will then use R to calculate the contingency table and the number of patients in each group.

```
survey.table<- table(survey.data$Receive.Letter,
                     survey.data$Return.Survey)
groups<- margin.table(survey.table, 1)
```

  Finally the calculation of the sample odds ratio confidence interval for the population odds ratio is:

```
or.ci(survey.table[,1], groups, conf.level = 0.95)

## Estimated Odds Ratio =  0.9462421
## Confidence Interval for Population Odds Ratio =  0.8749959 1.02329
```