2 x 2 Table Analysis

Patrick Kelly

2/20/2020

Click Here"Odds ratio – Confidence Interval"

What is a Contingency Table?

A contingency table summarises the outcomes of each individual sampled in terms of whether Properties (A - Exposure) and (B - Outcome) are absent or present. It represents the joint frequency distribution of the two properties.

		OUTCOME	
		Present	Absent
EXPOSURE	Present	a	b
	Absent	c	d

Data from case-control studies (retrospective or prospecitve) can be analyzed in several ways.

Odds Ratio

An odds ratio is a measure of association between the presence or absence of two properties.

Smoking and Cancer

In 1950, the Medical Research Council conducted a case-control study of smoking and lung cancer (Doll and Hill 1950).

Let's create a 2 X 2 table of the results.

```r
CT <- matrix(c(647, 2, 622, 27), nrow = 2)
rownames(CT) <- c("Smoker", "Non-Smoker")
colnames(CT) <- c("Cancer", "No-Cancer")
CT
```

```
##            Cancer No-Cancer
## Smoker        647       622
## Non-Smoker      2        27
```

Load R packages

```r
suppressMessages(library(oddsratio))
suppressMessages(library(questionr))
suppressMessages(library(DescTools))
# suppressMessages(library(epitools))
suppressMessages(library(fmsb))
```

Analysis

OR = ad/bc = (647 x 27) / (622 x 2)

```r
a <- CT[1]
c <- CT[2]
b <- CT[3]
d <- CT[4]
(647 * 27) / (622 * 2)
```

```
## [1] 14.0426
```

```r
OR <- round((a*d)/(b*c),2)
OR
```

```
## [1] 14.04
```

The odds of lung cancer in smokers is estimated to be 14 times the odds of lung cancer in non-smokers. How reliable is this estimate? We need to calculate a confidence interval. If the study is repeated and the range calculated each time, you would expect the true value to lie within these ranges on 95% of trials.

```r
OddsRatio(CT[1:2,], method="wald",
      conf.level=0.95)
```

```
## odds ratio     lwr.ci     upr.ci
##  14.042605   3.325329  59.300825
```

The 95% confidence interval for this odds ratio is between 3.33 and 59.3. Why such a huge range? It's because the numbers of non-smokers, particularly for lung cancer cases, are very small. Increasing the confidence level to 99% this interval would increase to between 2.11 and 93.25.

```
OddsRatio(CT[1:2,], method="wald",
      conf.level=0.99)
```

```
## odds ratio     lwr.ci     upr.ci
##  14.042605   2.114719  93.248662
```

Interpretation of case/control study

Patients with cancer or 14 times more likely to have been smokers than non-smokers.

Details of the CI algorithm

```
log_OR <-log((a*d)/(b*c))
log_OR
```

```
## [1] 2.642096
```

```
std_log_OR <- sqrt(1/a + 1/b + 1/c +1/d)
std_log_OR
```

```
## [1] 0.7349764
```

```
# Two tailed Z = 1.96, alpha = 0.05
ci_ll <-round(exp(log_OR - 1.96 * std_log_OR),2)
# ci_ll
ci_ul <- round(exp(log_OR + 1.96 * std_log_OR),2)
# ci_ul
cat("The 95% CI ranges from",ci_ll,"to",ci_ul)
```

```
## The 95% CI ranges from 3.33 to 59.3
```

Relative Risk

RR = a/(a+b) / c(/c+d)

```
RR <- (a/(a+b)) / (c/(c+d))
round(RR,2)
```

```
## [1] 7.39
```

```
fmsb_RR <- riskratio(647, 2, 1269, 29, conf.level=0.95, p.calc.by.independence=TRUE)
```

```
##            Disease Nondisease Total
## Exposed        647        622  1269
## Nonexposed       2         27    29
```

```
round(fmsb_RR$estimate,2)
```

```
## [1] 7.39
```

```
round(fmsb_RR$conf.int,2)
```

```
## [1]  1.94 28.19
## attr(,"conf.level")
## [1] 0.95
```

Interpretation of the RR

We are 95% confident that the relative risk of cancer in smokers compared to non-smokers is between 1.91 amd 28.19. The null value is 1. Since the 95% confidence interval does not include the null value (RR=1), the finding is statistically significant.

Another study

Does chocolate consumtion reduce the risk of cardiovascular disease?

Odds Ratio

```
CT2 <- matrix(c(925, 1020, 168, 147), nrow = 2)
rownames(CT2) <- c("Chocolate","None")
colnames(CT2) <- c("CV Disease","No-CV Disease")
CT2
```

```
##           CV Disease No-CV Disease
## Chocolate        925           168
## None            1020           147
```

```
odds.ratio(CT2)
```

```
##                      OR   2.5 % 97.5 %       p
## Fisher's test 0.79359 0.62032 1.0144 0.05969 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The OR at 0.79 suggests that chocolate (1-3 times a month) has some protective effect. But since the 95% confidence interval includes the null value of 1, the effect is not statistically significant.(p>0.05)

What about chocolate more than 4 times a week?

```r
CT3<- matrix(c(43, 168, 736, 925), nrow = 2)
rownames(CT3) <- c("Choc","None")
colnames(CT3) <- c("CV Disease","No-CV Disease")
CT3
```

```
##      CV Disease No-CV Disease
## Choc         43           736
## None        168           925
```

```r
odds.ratio(CT3)
```

```
##                     OR   2.5 % 97.5 %          p
## Fisher's test 0.32183 0.22155 0.4593 7.516e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The OR at 0.32 suggests that chocolate (more than 4 times a week) has a protective effect.The 95% confidence interval does not include the null value of 1, thus the effect is statistically significant. The risk reduction = 1 - OR = 0.68 = 68%. ($p<0$)

What abbut the Relative Risk?

```r
RRC <- (43/(43+736)) / (168/(168+925))
RRC
```

```
## [1] 0.3591219
```

```r
fmsb_RR2 <- riskratio(43, 168, 779, 1093, conf.level=0.95, p.calc.by.independence=TRUE)
```

```
##            Disease Nondisease Total
## Exposed         43        736   779
## Nonexposed     168        925  1093
```

```r
round(fmsb_RR2$estimate,2)
```

```
## [1] 0.36
```

```r
round(fmsb_RR2$conf.int,2)
```

```
## [1] 0.26 0.50
## attr(,"conf.level")
## [1] 0.95
```

```r
Percent_decrease <- (1 - RRC) * 100
Percent_decrease
```

```
## [1] 64.08781
```

Interpretation

Those who ate chocolate more than 4 times a week have 0.36 times the risk of cardiovascular disease compared to those who didn't eat chocolate. Since the 95% confidence interval did not include 1, the result is statistically signiicant.

Chocolate eaters had a cumulative incidence of CV disease of $43/779 = 0.055$ compared to $168/1093 = 0.154$ for non-chocolate eaters.

The chocolate eaters had a 64% decrease in CV disease risk.

What about a Chi Squared test?

Click Here"Chi Squared Test"

Hypotheses of variabe independence

H0: The 2 variables are independent

HA: They are related

Do the test without Yates correcton.

```
chisq.test(CT, correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  CT
## X-squared = 22.044, df = 1, p-value = 2.664e-06
```

```
# Reject H0 (p<0.05)
chisq.test(CT2, correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  CT2
## X-squared = 3.621, df = 1, p-value = 0.05706
```

```
# Do not reject H0 (p>0.05)
chisq.test(CT3, correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  CT3
## X-squared = 44.131, df = 1, p-value = 3.072e-11
```

```
# Reject H0 (p<0.05)
```

2 x 2 Classification Table

Resulting from Logistic Regression, for example.

The four data counts represent true and false positives and true and false negatives.The analysis is done with a confusion matrix which provides many statistics including: total accuracy, sensitiviy, specificity, precsion, recall and F1-Score. And then one can proede to the Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) statistics.

Click Here“Confidence Intervals”

Click Here“Technical Papers”

Click Here“Setup MAC to wirte technical papers”

Click Here“Odds Ratio”

Click Here“Odds Raatios Mislead”

Click Here“Association”

Click Here“Confusion Matrix”