Lab3

Ziwen Zhang

2022-11-17

# Question 1

# create a matrix with aspirin values  
aspirin <- matrix(c(189, 104, 10845, 10933), nrow = 2)  
aspirin

## [,1] [,2]  
## [1,] 189 10845  
## [2,] 104 10933

# write a function to calculate odds ratio, relative risk ratio and difference of proportions for a 2 x 2 table.  
calculate <- function(data){  
 odds1 <- data[1,1] / data[1,2]  
 odds2 <- data[2,1] / data[2,2]  
 odds\_ratio <- odds1 / odds2  
 low\_OR <- exp((ln(odds\_ratio) - 1.96 \* sqrt(1/data[1,1] + 1/data[1,2] + 1/data[2,1] + 1/data[2,2])))  
 upp\_OR <- exp((ln(odds\_ratio) + 1.96 \* sqrt(1/data[1,1] + 1/data[1,2] + 1/data[2,1] + 1/data[2,2])))  
   
 relative\_risk1 <- data[1,1]/(data[1,1] + data[1,2])  
 relative\_risk2 <- data[2,1]/(data[2,1] + data[2,2])  
 relative\_risk\_ratio <- relative\_risk1 / relative\_risk2  
 low\_RR <- exp((ln(relative\_risk\_ratio) - 1.96 \* sqrt(1/data[1,1] + 1/data[2,1] - 1/(data[1,1] + data[1,2]) - 1/(data[2,1] + data[2,2]))))  
 upp\_RR <- exp((ln(relative\_risk\_ratio) + 1.96 \* sqrt(1/data[1,1] + 1/data[2,1] - 1/(data[1,1] + data[1,2]) - 1/(data[2,1] + data[2,2]))))  
   
 diffProp <- data[1,1]/(data[1,1] + data[1,2]) - data[2,1]/(data[2,1] + data[2,2])  
 pi1 <- data[1,1]/(data[1,1] + data[1,2])  
 pi2 <- data[2,1]/(data[2,1] + data[2,2])  
 low\_DP <- diffProp - 1.96 \* sqrt((pi1\*(1-pi1))/(data[1,1]+data[1,2]) + (pi2\*(1-pi2))/(data[2,1]+data[2,2]))  
 upp\_DP <- diffProp + 1.96 \* sqrt((pi1\*(1-pi1))/(data[1,1]+data[1,2]) + (pi2\*(1-pi2))/(data[2,1]+data[2,2]))  
   
 paste("95% CI for odds ratio is from ", round(low\_OR,3), "to", round(upp\_OR,3), "95% CI for relative risk ratio is from ", round(low\_RR,3), "to", round(upp\_RR,3), "95% CI for difference of proportions is from ", round(low\_DP,3), "to", round(upp\_DP,3))  
}  
  
# apply the function on aspirin data  
calculate(aspirin)

## [1] "95% CI for odds ratio is from 1.44 to 2.331 95% CI for relative risk ratio is from 1.433 to 2.306 95% CI for difference of proportions is from 0.005 to 0.011"

Comments: The 95% CI for odds ratio is a little bit greater than that for the risk ratio. This is because the odds ratio overestimates the risk if the disease is common. (when the disease is rare, the odds ratio and relative risk will be similar.)

# Question 2

party\_tab = matrix(c(762, 484, 327, 239, 468, 477), nrow=2)  
  
calculate2 <- function(dt){  
 party\_chisq <- chisq.test(dt)  
 obs <- party\_chisq$observed  
 exp <- party\_chisq$expected  
 Gscore <- 2 \* sum(obs \* log(obs/exp))  
 pvalue <- pchisq(q = Gscore, df = (nrow(dt)) \* (ncol(dt)-1), lower.tail = FALSE)  
 paste("The G2 acore is: ", round(Gscore,3), "and the p-value is: ", signif(pvalue,3))  
}  
  
calculate2(party\_tab)

## [1] "The G2 acore is: 30.017 and the p-value is: 4.86e-06"

# Question 3

result <- matrix(c(7, 0, 8, 15), nrow=2)  
dimnames(result) <- list(Predniosolone = c("Treated","Control"), Normalization = c("achieved", "not achieved"))  
result

## Normalization  
## Predniosolone achieved not achieved  
## Treated 7 8  
## Control 0 15

fisher.test(result, alternative = "greater")

##   
## Fisher's Exact Test for Count Data  
##   
## data: result  
## p-value = 0.003161  
## alternative hypothesis: true odds ratio is greater than 1  
## 95 percent confidence interval:  
## 2.645931 Inf  
## sample estimates:  
## odds ratio   
## Inf

Since the p-value is smaller than 0.05, there is an evidence to reject the null hypothesis and thus the results were significantly better for treatment than for control.

# Question 4

score <- matrix(c(9, 44, 13, 10, 11, 52, 23, 22, 9, 41, 12, 27), ncol = 3)  
dimnames(score) <- list(Aspiration = c("High School","High School Graduation", "Some College", "College Graduation"), Income = c("Low","Middle", "High"))  
score

## Income  
## Aspiration Low Middle High  
## High School 9 11 9  
## High School Graduation 44 52 41  
## Some College 13 23 12  
## College Graduation 10 22 27

# using Chi-squard test to test independence of educational aspiration and family income.  
score\_chisq <- chisq.test(score)  
score\_chisq

##   
## Pearson's Chi-squared test  
##   
## data: score  
## X-squared = 8.8709, df = 6, p-value = 0.181

Since the p-value is 0.181, there is no evidence to reject the null hypothesis that educational aspiration and family income are independent.

Deficiency of Chi-square test: The Chi-square treat X and Y as nominal, reordering does not change the chi-square value, but it is not very powerful when testing the ordinal variables.

library(DescTools)

## Warning: package 'DescTools' was built under R version 4.2.2

MHChisqTest(score)

##   
## Mantel-Haenszel Chi-Square  
##   
## data: score  
## X-squared = 4.7489, df = 1, p-value = 0.02932

Since the p-value is smaller than 0.05, there is an evidence to reject the null hypothesis that the educational aspirations and family income are independent.