Lab2

Ziwen Zhang

2022-11-10

# Question 1

# read the file  
arthritis <- read.csv("arthritis.csv")  
  
# reordered the "Improved" column  
Improvedord <- ordered(arthritis$Improved, levels=c("None", "Some", "Marked"))  
  
# create a new data frame with the reordered Improved column  
new <- data.frame(arthritis, Improvedord)  
  
# make a 3-way table  
mytable <- xtabs(~ Treatment+Sex+Improvedord, data=arthritis)  
ftable(mytable)

## Improvedord None Some Marked  
## Treatment Sex   
## Placebo Female 19 7 6  
## Male 10 0 1  
## Treated Female 6 5 16  
## Male 7 2 5

# Question 2

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

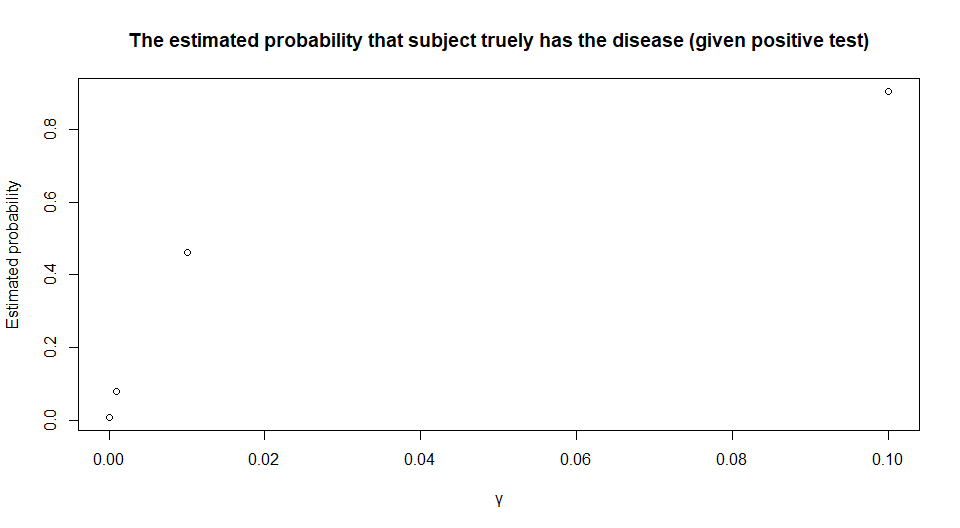
## [,1] [,2]  
## [1,] 189 10845  
## [2,] 104 10933

# 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  
   
 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  
   
 diffProp <- data[1,1]/(data[1,1] + data[1,2]) - data[2,1]/(data[2,1] + data[2,2])  
   
 paste("odds ratio: ", round(odds\_ratio,3), "relative risk ratio: ", round(relative\_risk\_ratio,3), "difference of proportions: ", round(diffProp,3))  
}  
  
# apply the function on aspirin data  
calculate(aspirin)

## [1] "odds ratio: 1.832 relative risk ratio: 1.818 difference of proportions: 0.008"

# Question 3

# set gama values  
gama <- c(0.1, 0.01, 0.001, 0.0001)  
  
# calculate estimated probability  
pEst <- 0.85\*gama / (0.85\*gama + 0.01\*(1-gama))  
  
# create a data frame to combine gama values and the results  
result <- data.frame(gama, pEst)  
  
# plot the result  
plot(result$gama, result$pEst,   
 main = "The estimated probability that subject truely has the disease (given positive test)",   
 xlab = "γ",   
 ylab = "Estimated probability")



As the probability of a subject has the disease (γ values) increases, the estimated probability that subject truly has the disease (given positive test) increases as well. The trend is more significant when γ values are small.

# Question 4

Consider an emergency room that, on the average, accepts 5 patients per 3 hours. • What is the probability that there will be no patients next hour? • At least two patients next hour?

# set Y as the number of patients per hour, then Y has Poisson distribution with λ = 5/3  
  
prob1 <- dpois(0, 5/3, log=FALSE)  
paste("The probability that there will be no patients next hour is ", round(prob1, 4))

## [1] "The probability that there will be no patients next hour is 0.1889"

prob2 <- ppois(2-1, 5/3, lower.tail = FALSE, log.p = FALSE)  
paste("The probability that at least two patients next hour is ", round(prob2, 4))

## [1] "The probability that at least two patients next hour is 0.4963"