Problem Set 2

applications of probability theory

C. Durso

# Introduction

These questions were rendered in R markdown through RStudio (<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>, <http://rmarkdown.rstudio.com> ).

Please generate your solutions in R markdown and upload both a knitted doc, docx, or pdf document in addition to the Rmd file. Please put your name in the “author” section in the header.

The questions in this problem set use material from the slides on discrete and continuous probability spaces and the Rmds Discrete\_Probability\_Distributions\_2\_3\_3.Rmd and 02\_continuous\_probability\_distributions\_in\_r.rmd

# Load Data

data("PolioTrials")  
dat<-PolioTrials

# Question 1

Please carry out the analysis below and answer the questions that follow. For this assignment, please do all calculations in R and show the code and the results in the knit document.

## Context

Question 2 on problem set 1 addresses the question of whether the NotInoculated and Placebo groups in the Randomized Control experiment had statistically significantly different rates of paralytic polio.

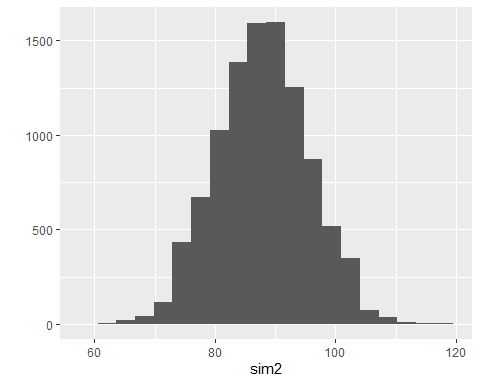
Recall that the NotInoculated and Placebo groups differ in that the children in the Placebo group had been enrolled in the vaccine trial while the parents of the children in the NotInoculated group did not enroll their children.

The approach, using the rbinom function, implemented the idea that populations in the NotInoculated and Placebo groups in the RandomizedControl experiment were the same in regards to paralytic polio cases by using the rbinom function to assign paralytic polio cases in the combined NotInoculated and Placebo groups of the RandomizedControl experiment to the Placebo group with probability equal to the ratio of the size of the Placebo group to the size of pooled Placebo group and NotInoculated group.

Note that the function rbinom(x,size,prob) simulates drawing random samples from Binom(size,prob).

The computations for that analysis are reproduced here:

n<-10000 # number of simulations  
  
# Calculate the number of paralytic polio cases in the pooled "Placebo" and "NotInoculated" group.  
ct<-sum(dat$Paralytic[2:3])  
  
# Calculate the proportion "prop" of the the pooled "Placebo" and "NotInoculated" group that are in the "Placebo" group.  
prop<-dat$Population[2]/sum(dat$Population[2:3])  
  
# Generate 10,000 counts of paralytic polio cases in the "Placebo" group under the model that each paralytic polio case in the pooled pooled "Placebo" and "NotInoculated" group has probability "prop" of being assigned to the "Placebo" group.  
set.seed(45678765)  
sim2<-rbinom(n,ct,prop)   
qplot(sim2,bins=20)



# proportion of the simulated counts of paralytic polio in the "Placebo" that are less than or equal to the observed count:  
mean(sim2<=dat$Paralytic[2])

## [1] 0.9997

# proportion of the simulated counts of paralytic polio in the "Placebo" that are greater than or equal to the observed count:  
mean(sim2>=dat$Paralytic[2])

## [1] 4e-04

## Q1, part 1

(10 points)

Using the same null model described above, please calculate the probability that the count of paralytic polio cases in the Placebo group under the null model is less than or equal to dat$Paralytic[2] directly rather than by simulating it. Recall that the function pbinom(x,size,prob) returns the probability of the event that the number of successes is in the set .

(Your answer here

# Please be sure that your computed probability shows in your knitted solutions

)

## Q1, part 2

(10 points)

Using the same null model described above, please calculate the probability that the count of paralytic polio cases in the Placebo group under the null model is greater than or equal to dat$Paralytic[2] directly rather than by simulating it. Hint: Denote the value in part 1 by . This answer is not . The value is the probability of the event that count of paralytic polio cases in the Placebo group under the null model is strictly greater than dat$Paralytic[2].

(Your answer here

# Please be sure that your computed probability shows in your knitted solutions

)

## Q1, part 3

(10 points)

Is the value computed in part 2 strong evidence against the null model? (your yes or no answer here, with an explanation based on the calculations above)

# Question 2

## Context

This question concerns the uniform distribution on , the continuous probability space with and defined by for measurable sets as described in the week 2 slides. This distribution will be important in hypothesis testing.

## Q2

(10 points)

Are the events and independent. To answer this, please address the following questions:

1. What is ? (your answer here)
2. What is ? (your answer here)
3. What is ? (your answer here)
4. Are the events and independent? (your answer here: please answer yes or no and explain your response using the calculations above.)

# Question 3

## Context

This question concerns the standard Normal distributions, the continuous probability space with and defined by for measurable sets as described in the week 2 slides. This distribution will be essential in future methods.

## Q3

(10 points)

Are the events and independent. To answer this, please address the following questions:

1. Please give a numerical approximation to . (your answer here

)

1. Please give a numerical approximation to ? (your answer here

)

1. Please give a numerical approximation to . (your answer here

)

1. Are the events and independent? (your answer here: please answer yes or no and explain your response using the calculations above.)