Lab02\_Massey

Matt Massey

5/29/2021

Question 1.

#to generate 23 birthdays  
set.seed(1)  
dates <- sample(1:365, 23, replace = TRUE)  
  
#sort  
sorted <-sort(dates)  
sorted

## [1] 37 79 85 89 105 129 165 167 187 213 217 263 270 277 289 290 299 307 324  
## [20] 329 330 362 362

#check for duplicates so we can see them  
repeats <- rle((sort(dates)))$length  
repeats

## [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2

#this checks to see if there are duplicates  
length(dates) != length(unique(dates)) #TRUE if there are duplicates

## [1] TRUE

Question 1b. In the simulation above (with set.seed(1)), two people have the same birthday on day 362 (December 28). The other 21 people in the above simulation have unique birthdays. There are not three people who have the same birthday.

Questions 2-4.

#repeat the above sample 10,000 times  
set.seed(1) # not mentioned in lab 2 instructions  
dupe\_count = 0 # set counter to count number of repeat birthdays  
runs =10000  
for (i in 1:runs) {  
 dates <- sample(1:365, 23, replace = TRUE)  
 if (length(dates) != length(unique(dates))) {  
 dupe\_count = dupe\_count + 1  
 }  
}  
  
#printing the percentage of duplicate birthdays to est. probability  
print(dupe\_count/runs)

## [1] 0.5026

*\*Note: the code from lab 2 (above) does not match the instructions. The instructions call for 1,000 repeats of the simulation, but this code is for 10,000. Also, the last line of Question 2 needs edited (it mentions samples of 50). Also, the video uses set.seed(1), which is not mentioned in the lab instructions for questions 2-4, nor in the code, but I did use it here.*

Question 2. It can be estimated that the probability of at least two people in a group of 23 having the same birthday is 0.5026.

Question 3. Yes, this is completely surprising. It is not intuitive that at least two people in a group of 23 would be ~50% likely to share the same birthday.

Question 4. Wikipedia shows an in-depth explanation of the ‘birthday problem’ and states that about 23 people (or fewer) are needed to reach the 0.5 probability threshold that two of those people will share the same birthday. This confirms the simulation from r and I still find counterintuitive, however, the explanation that the problem is actually looking at all possible pairs of people, with a group of 23 you have 253 pairs, which is over half the days of the year. Although that explanation helps, I’m still very surprised.