Lecture 1: Intro to simulation

Warm-up question

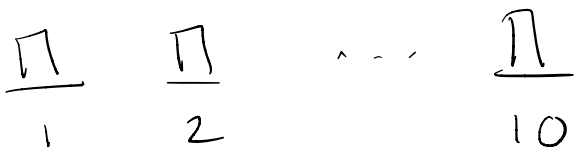
Problem: 10 people are at a party, and all of them are wearing hats. They each place their hat in a pile; when they leave, they choose a hat at random. What is the probability at least one person selected the correct hat?

Question: Work with your neighbor to discuss the following question:

- Without calculating probabilities, how could you design an experiment to estimate this probability?

Designing an experiment

Step 1: need 10 hats!

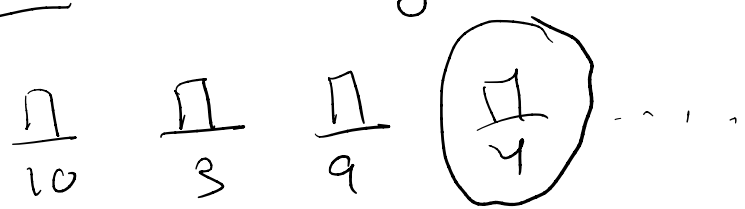


← need to represent hats in R

Step 2: randomly assign hats

← need to "shuffle" hats in R

Step 3: who got their original hat?



Step 4: Repeat many times!

← repeat process in R

Step 1: representing the hats

```
1 hats <- 1:10
2
3 hats
```

hats <- 1:10 ← vector contains 1, 2, 3, ..., 10

store vector as a variable named "hats"

```
 [1]  1  2  3  4  5  6  7  8  9 10
```

```
1 hats[3]
```

```
[1] 3
```

- `hats` is a **vector**, containing the numbers 1 to 10
- entries in a vector are accessed by their index

Step 2: everyone draws a random hat

```
1  hats <- 1:10
2  randomized_hats <- sample(hats, size = 10, replace = FALSE)
3
4  hats
```

# of Samples

what we sample from

Sample without replacement

```
 [1]  1  2  3  4  5  6  7  8  9 10
```

```
1  randomized_hats
```

```
 [1]  8  2  4  5  9  7  1 10  3  6
```

received correct hat!

- The `sample` function creates a random sample from a vector
- How many people selected their original hat?

Step 3: check who got their original hat

```
hats <- 1:10
randomized_hats <- sample(hats, size = 10, replace = FALSE)
```

```
hats
```

```
 [1]  1  2  3  4  5  6  7  8  9 10
```

```
randomized_hats
```

```
 [1]  8  2  4  5  9  7  1 10  3  6
```

```
hats == randomized_hats
```

== test for equality

```
 [1] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
# TRUE is 1, FALSE is 0
sum(hats == randomized_hats)
```

how many people got their original hat?

```
[1] 1
```

```
# did at least one person get their hat?
sum(hats == randomized_hats) > 0
```

```
[1] TRUE
```

(at least one person received their original hat)

Code so far

```
1 hats <- 1:10
2 randomized_hats <- sample(hats, size = 10, replace = FALSE)
```

```
1 sum(hats == randomized_hats) > 0
```

```
[1] TRUE
```

- Is this a good estimate of the probability?

No! Need to repeat many times

Step 4: iteration

A `for` loop repeats code many times:

```
1 nsim <- 10000 # number of simulations
2 for(i in 1:nsim){
3 
4 
5 }
```

"repeat the following code for $i = 1, 2, 3, \ldots, nsim$"

Step 4: iteration

A `for` loop repeats code many times:

```
1  nsim <- 10000 # number of simulations
2  hats <- 1:10
3  results <- rep(NA, nsim) # vector to store results
4
5  for(i in 1:nsim){
6    randomized_hats <- sample(hats, size = 10, replace = FALSE)
7    results[i] <- sum(hats == randomized_hats) > 0
8  }
9
10 head(results)
```

```
[1] FALSE FALSE  TRUE FALSE FALSE FALSE
```

repeat NA nsim times

NA, NA, NA, ...

randomly assign hats

Store the result

check whether at least one person got their original hat

want proportion of times result was TRUE

Step 4: iteration

A `for` loop repeats code many times:

```
1  nsim <- 10000 # number of simulations
2  hats <- 1:10
3  results <- rep(NA, nsim) # vector to store results
4
5  for(i in 1:nsim){
6    randomized_hats <- sample(hats, size = 10, replace = FALSE)
7    results[i] <- sum(hats == randomized_hats) > 0
8  }
9
10 mean(results)
```

magic number!

magic number!

```
[1] 0.6373
```

P(at least one person receives original hat) ≈ 0.637

- What if I wanted to repeat the simulation, with a different number of people?

magic number: value with no clear meaning

Removing magic numbers

Without magic numbers:

```
 1  nsim <- 10000 # number of simulations
 2  M <- 10 # number of people
 3  hats <- 1:M
 4  results <- rep(NA, nsim) # vector to store results
 5
 6  for(i in 1:nsim){
 7    randomized_hats <- sample(hats,
 8                              size = M,
 9                              replace = FALSE)
10    results[i] <- sum(hats ==
11                        randomized_hats) > 0
12  }
13
14  mean(results)
```

```
[1] 0.6285
```

- Why did I get different results?

Final code

```r
set.seed(3) # set a seed for reproducibility

M <- 10 # number of people at the party
hats <- 1:M # numbered hats
nsim <- 10000 # number of simulations
results <- rep(NA, nsim) # vector to store the results

for(i in 1:nsim){
  # hats are randomly assigned to each person
  randomized_hats <- sample(hats, M, replace = F)

  # did at least one person get their hat back?
  results[i] <- sum(randomized_hats == hats) > 0
}

mean(results)
```

seed

results = NA, NA, ..., NA (10000 entries)

comments

ith entry of results

Start: results NA, NA, NA, ...

i = 1 results TRUE, NA, NA, ...

i = 2 results TRUE, FALSE, NA, ...

Summary of coding practices

- avoid magic numbers
- set a seed for reproducibility
- use meaningful names
- add comments

Class activity

Work with a neighbor on the class activity (link below and on the course website):

https://sta279-f23.github.io/class_activities/ca_lecture_1.html