

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: representing the hats

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

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

```
1 randomized_hats
```

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

- 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

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

```
1 hats
```

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

```
1 randomized_hats
```

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

```
1 hats == randomized_hats
```

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

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

```
[1] 2
```

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

```
[1] TRUE
```

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?

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 }
```


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 TRUE TRUE FALSE TRUE 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)
```

```
[1] 0.6357
```

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

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] 0.6269
```

- Why did I get different results?

Final code

```
1  set.seed(3) # set a seed for reproducibility
2
3  M <- 10 # number of people at the party
4  hats <- 1:M # numbered hats
5  nsim <- 10000 # number of simulations
6  results <- rep(NA, nsim) # vector to store the results
7
8  for(i in 1:nsim){
9    # hats are randomly assigned to each person
10    randomized_hats <- sample(hats, M, replace = F)
11
12    # did at least one person get their hat back?
13    results[i] <- sum(randomized_hats == hats) > 0
14  }
15
16  mean(results)
```

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-s24.github.io/class_activities/ca_lecture_1.html

