Problem Set 1

Ray Hossain

2023-09-24

1. Conducting a survey. (12 points, 2 points per question) Suppose the Hertie School has 600 affiliated members: 400 students, 50 faculty, and 150 staff. An administrator hires you to conduct a survey of the Hertie School community to learn what can be improved.

    a. Unfortunately, you haven't been given the budget to survey everyone, so you have to randomly sample 60 individuals. Also, for some reason you've been instructed to sample with replacement. What is the probability that you sample 60 unique respondents? In other words, what is the probability that, when you sample with replacement, you don't give the survey to the same person more than once?

```
n <- 600
matches <- rep(NA, n)

for (i in 1:n) {
  respondents <- sample(1:600, 60, replace = TRUE)
  matches[i] <- sum(duplicated(respondents))
}

prop.table(table(matches == 0))
```

```
##
## FALSE  TRUE
##  0.96  0.04
```

```
##This is similar to the formula answer.

# Define the population size and the sample size
k <- 60

# Calculate the probability of sampling 60 unique respondents with replacement
p_unique <- prod(n:(n-k+1)) / n^k

# Print the result
print(round(p_unique*100, 2))
```

```
## [1] 4.72
```

</ There is only a 4 percent chance that you do not give the survey to the same person more than once.

    b. You present the administrator with the answer to part (a) and convince them that it would really make more sense to sample without replacement. How many ways are there to construct your sample? (We don't care about order here — just the composition of the group.

```r
choose(600, 60)
```

```
## [1] 2.774267e+83
```

c. What is the probability that the survey sample contains only students?

```r
# Define the population size, the sample size, and the number of students
k <- 60
s <- 400

# Calculate the probability of sampling only students without replacement
p_students <- choose(s, k) / choose(n, k)

# Print the result
print(p_students)
```

```
## [1] 5.438102e-12
```

d. To guard against this (admittedly distant) possibility, let's construct our sample in such a way that ensures equal representation: we randomly sample exactly 40 students from the 400, then 5 faculty from the 50, and 15 staff from the 150. This technique is called blocking. How many ways are there to construct the survey sample now? (Again, we don't care about order here.)

```r
# Define the population size and the sample size for each group
n_students <- 400
n_faculty <- 50
n_staff <- 150
k_students <- 40
k_faculty <- 5
k_staff <- 15

# Calculate the number of ways to construct the survey sample with blocking
n_ways <- choose(n_students, k_students) * choose(n_faculty, k_faculty) * choose(n_staff, k_staff)

# Print the result
print(n_ways)
```

```
## [1] 6.779343e+81
```

e. Now, you want to follow up on your survey with a focus group of 10 people, also randomly sampled from the 600, who will be arranged around a circular table. How many ways can you construct your focus group when they have assigned seats? (In other words, we count the same group of people in two different seating arrangements as two distinct ways to construct the focus group.)

```r
k_focus <- 10

arrangement <- choose(n, k_focus) * factorial(k_focus)

print(arrangement)
```

```
## [1] 5.607472e+27
```

f. Now consider only the focus group samples that contain five students and five non-students. Within this subset, what is the probability that all five students are seated next to each other?

```
(factorial(5) * choose(400, 5) * choose(200, 5)) / (choose(600, 10))
```

```
## [1] 16.38652
```