Sample & Population

Code authors: Jinhui Li, Viviana Monje.

Corresponding author: vmonje@buffalo.edu

Recommended Matlab: >= 2023a.

Table of Contents

Questionnaire prior to use of this NB

[These questions will be presented to students as a poll the week before introducing the notebook in class]

- What concepts from statistics do you think you will use the most in your engineering career?
- Did you previously take a course in statistics?
- Provide a concrete example of a sample.
- Provide a concrete example of a population.
- How would you define the concept of 'probability'?

Usage Rights

These materials were created for educational research purposes and to be completed as part of material covered in the CE305: Probability, Statistics, and Data Analysis undergraduate course. Students can benefit from the content here presented whether they consent to participate in the research study or not.

Learning Objectives

[Before using this notebook in class, students have completed the reading assignment of the week, basic definitions in the sections below serve as quick reminders and include the corresponding textbook section]

By the end of this notebook, students will be able to:

- Define and identify sample and population;
- Utilize descriptive statistics to compute metrics for sample data;
- Identify sample space and its events
- Apply axioms of probablity to perform operations between events;
- Recognize the nature of experiments and use the appropriate counting method to determine expected number of outcomes.
- Compute probabilities of events for a known sample space.

Sample and Population

Basics

Population: the entire collection of objects or outcomes about which information is sought.

Sample: a subset of a population, containing the objects or outcomes that are actually observed.

Random experiment: an experiment where possible outcomes are known, but it is not possible to predict the exact outcome that will result individual times the experiment is run.

Detailed discussion and explanations are found in the Textbook, Ch. 1.1 Sampling. (Navidi Textbook, 6th Ed)
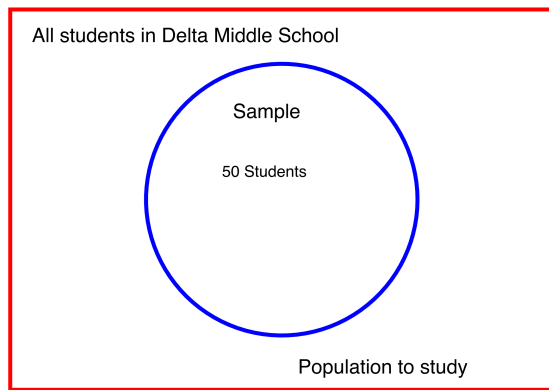
Example

You are a data analyst for a school-wide health project; you are responsible for collecting data on sleeping habits and quality of all students in the Delta Road Middle School. Your data will be analyzed to provide appropriate support for mental health and academic performance. You conduct a survey to estimate the average daily hours of sleep per student.

- Question 1: Who constitute the population of this study?

```
All students in the Delta Middle School.
```

- Question 2: Suppose you select 50 representative answers to report in the study, what is this subset of answers called?
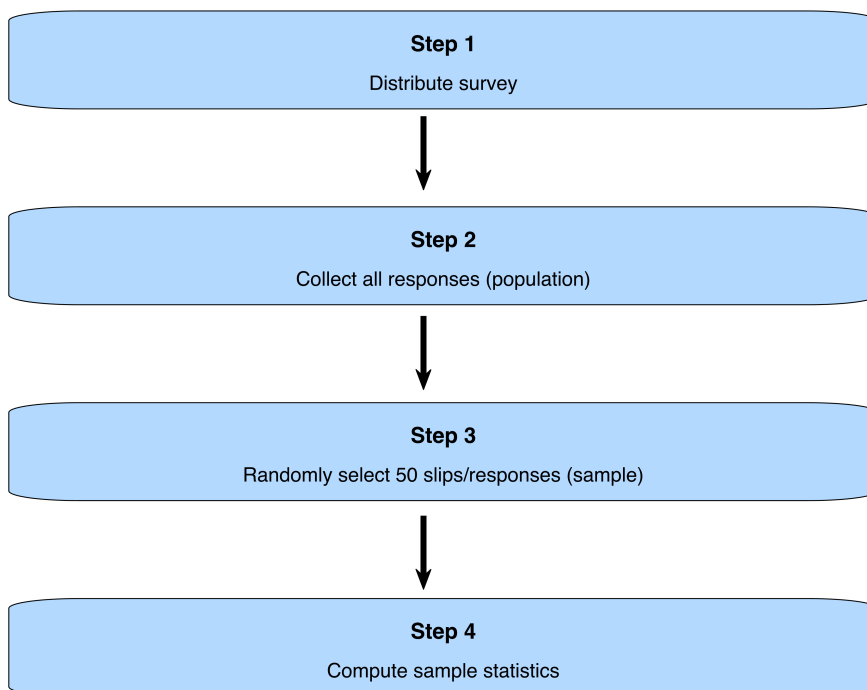
```
This subset of the population is called a sample.
```

All students in Delta Middle School

Sample

50 Students

Population to study

- Question 3: Given your question is: "How many hours do you spleep per night during the school year," is this a random experiment?

Yes, we know potential answers to the question, but not the number of hours each student will report.

- Question 4: How would you design the random experiment above?

Step 1

Distribute survey

Step 2

Collect all responses (population)

Step 3

Randomly select 50 slips/responses (sample)

Step 4

Compute sample statistics

- Question 5: If you repeat step 3 in the above diagram multiple times, selecting random answers each time, will you compute the same sample statistics?

```
No, sample statistics will change for every sample even thought they all come from the same
population. Sample statistics are point estimates of the population parameters
```

Descriptive Statistics

Basics

Let X_1, X_2, ..., X_3 be a sample:

- Sample Mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

- Sample Variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

- Sample Standard Deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$

More details can be found in Ch. 1.2 Summary Statistics. (Navidi Textbook, 6th Ed)

Example

Students in the statistics class take their first quiz; five (5) students selected at random have the following scores: {5, 8, 2, 5, 9}, then:

```
The current score list is:

5 8 2 5 9 10
```

- Sample Mean is:

```
The average is   6.50.
```

- Sample Variance is:

```
The sample variance is   9.10.
```

- Sample Standard Deviation:

```
The standard deviation is   3.02.
```
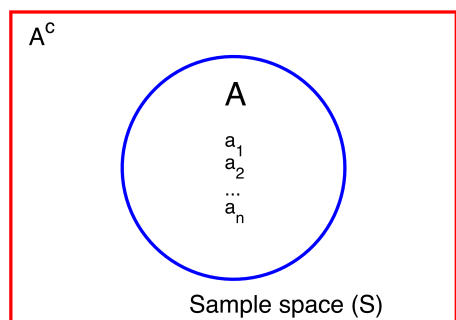
4

Events

Basics

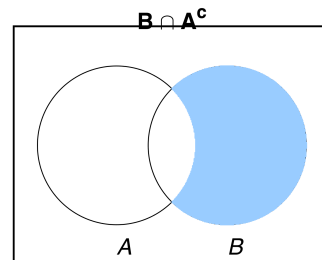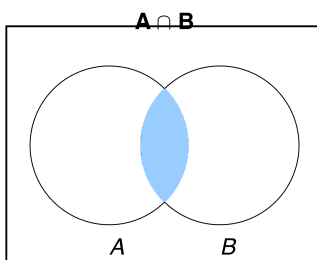Sample Space: the set of all possible outcomes of a random experiment.
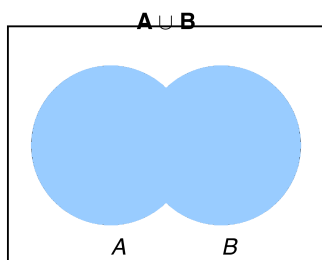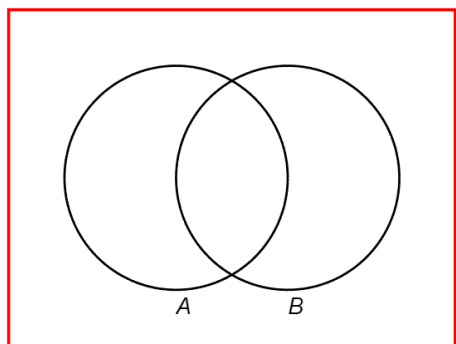
Event:

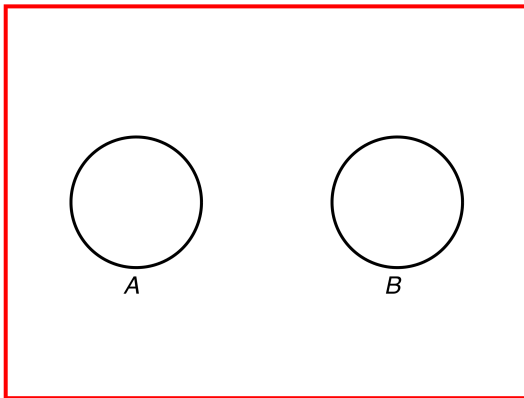More details can be found in Ch. 2.1 Basics Ideas. (Navidi Textbook, 6th Ed)

Event: A subset of a sample space.

Sample space (S)

Event Operations: Union, Intersection, and Complement.

Mutually Exclusive Events

The Axioms of Probability

1. Let S denote a sample space, the total probability of S is always: P(S) = 1.
2. For any event A in this sample space, 0 ≤ P(A) ≤ 1.
3. Individual outcomes in S are equally likely; the probability of an event A is given by the sum of probabilities of its individual outcomes: P(A) = P(a_1) + P(a_2) +…
4. If A and B are mutually exclusive events, then P(A ∪ B) = P(A) + P(B).
5. If A_1, A_2, … are mutually exclusive events, then P(A_1 ∪ A_2 ∪…) = P(A_1) + P(A_2) +…

More details can be found in Ch. 2.1 Basics Ideas. (Navidi Textbook, 6th Ed)

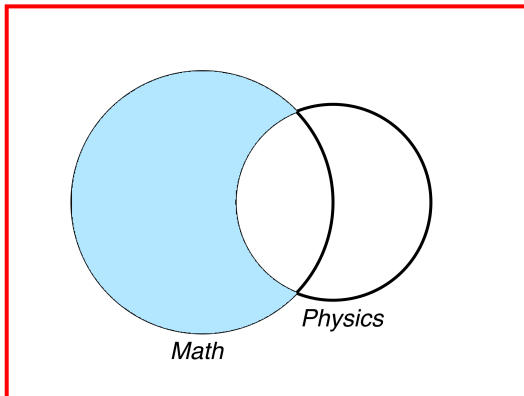## Example

A school has 100 enrolled students this year, 60 of them are enrolled in mathematics, 45 in physics, and 25 in both.

Questions:

- What is the size of the sample space?
- What are the events defined for this sample space?
- What is the probability students like math OR physics?
- What is the probability students do not math NOR physics?
- What is the probability students only like Math?

The probability of the event "only like math"is: 0.40

The probability of the event "like neither math nor physics" is: 0.15

The probability of the event "only like math"is: 0.40

Counting Methods

Basics

Probability: it is the number of times an outcome occurs out of the total possible number of outcomes in a sample space
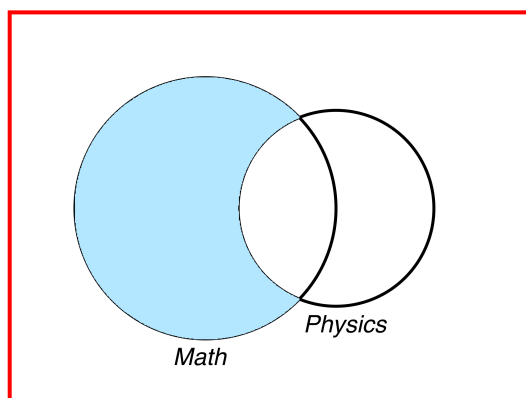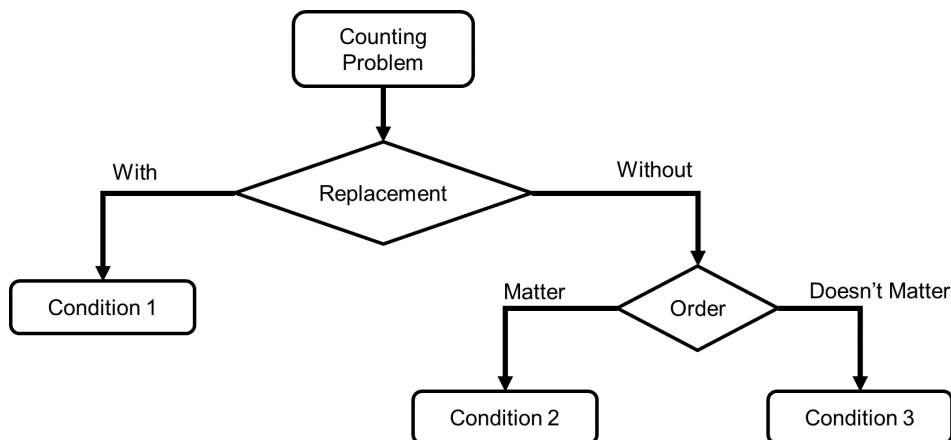
To compute probabilities of events in a sample space, we need efficient ways to count the number of outcomes in the sample space as well as for the event of interest. To this end, we need to identify the nature of the experiment and select the appropriate counting method.

More details can be found in Ch. 2.2 Counting Methods (Navidi Textbook, 6th Ed)

Counting events with replacement (order matters)

$$N = n^r$$

Counting events without replacement (order matters): Permutation.

$$N = \frac{n!}{(n-r)!}$$

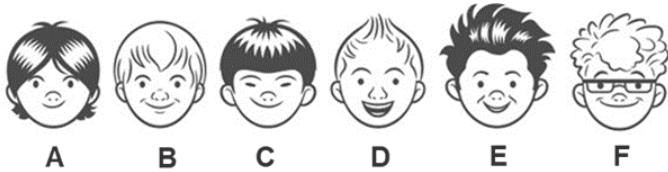Counting events without replacement (order doesn't matter): Combination.

$$N = \frac{n!}{r!(n-r)!}$$

Counting events from multiple groups (Distinguishable elements):

$$N = \frac{n!}{r_1! r_2! r_3! \cdots}$$

Example

There are six students in a class, labeled A, B, C, D, E, and F. The teacher plans to select students to participate in upcoming school sports events: **football** and **basketball**.

Situation 1: The teacher will select two different students to participate in football and basketball, respectively. How many options are there?

```
Situation 1: Two students are selected without replacement.

Each is assigned to football or basketball (roles matter).
Total options = P(6, 2) = 30
```

Situation 2: The teacher will select three different students to form a team and participate in a football game. How many options are there?

```
Situation 2: Three students are selected as a football team.
Order does not matter. The group is unordered.
Total options = C(6, 3) = 20
```

Situation 3: The teacher will select three different students. Among them, two students will participate in football, and one student will participate in basketball. How many options are there?

```
Situation 3: Three students are selected without replacement.
Two are assigned to football, one to basketball.
Roles matter between groups (football vs basketball),
but not within the football pair.
Total options = C(6,3) × C(3,2) = 20 × 3 = 60
```

Practice Activity After Class

[This portion of the notebook is assign to the students & graded for completion (not correctness). Students also have a regular problem set for the week for consistency during the semester]

Practice Activity 1: True or False

1. We want to understand the average study duration of all students in the school, so we survey one Grade 12 class. These surveyed students make up the population.

2. If we use a random number table to select 50 students from the entire school registry, this constitutes a simple random sample.

3. A sample must be larger than the population in order to produce accurate conclusions.

4. The population refers to the entire group of interest, while a sample refers to the subset of data used to study trends.

5. Surveying your 10 closest friends on their favorite course this semester is a simple random sampling method.

Practice Activity 2: Statistical analysis of napping times

You are part of your school's *"Healthy Detective"* project. Your mission is to investigate whether taking naps during the day helps improve students' learning performance. You randomly select 10 students and record their average nap duration per day (in hours) over the past week. Here is the data you collected:

Nap Duration Data (hours): 1.0, 0.5, 1.5, 2.0, 1.0, 0.0, 0.5, 3.0, 2.5, 1.5

- Calculate the mean nap duration.

- Determine the median nap duration.

- Calculate the sample variance of nap duration.

- Calculate the sample standard deviation of nap duration.

Practice Activity 3: Operations with Events

We have a prize box with the following items:

Number	Color	Prize Type
1	Red	Toy
2	Red	Candy
3	Blue	Toy
4	Blue	Toy
5	Yellow	Candy
6	Blue	Candy
7	Red	Toy
8	Yellow	Toy
9	Blue	Candy
10	Red	Candy

Questions:

- What is the size of sample space S?

- Given event A: "Drawing a red prize." How many elements does this event have?

- Given event B: "Getting a candy prize." How many elements does this event have?

- What is the probability P(A ∪ B)?

- What is the probability P(A ∩ B)?

- What is the probability P(B^C)?

- Are there any mutually exclusive events? If yes, please list; if no, why?

Practice Activity 4: Counting

Given the following tasks:

- Task A requires two people,
- Task B requires one person,
- Task C also requires one person.

You select four people out of ten to complete each a single task (no person completes more than one task).

Question:

How many different ways are there to assign these four people to complete the tasks?

Self-reflection Survey After Finishing the Notebook

[Upon submitting the assigned portion of the notebooks, the student will complete the following questionnaire that counts towards the grade for participation. Student responses will be compared to those of the initial questionnaire to determine improvements in clarity, quality, and sophistication of the responses. Concepts covered in this notebook will be evaluated one week after the intervention in the (regular) biweekly quiz]

1. Describe in your own words the difference between a sample and a population.
2. Provide an example of a random experiment.
3. How are counting methods used to compute probability of events in a sample space?
4. What aspects of this notebook were helpful or hindered your understanding of sample and population?