Sample & Population

Code authors: Jinhui Li, Viviana Monje.

Corresponding author: vmonje@buffalo.edu

Recommended Matlab: >= 2023a.

Table of Contents

Usage Rights

These materials were created for educational purposes and to be completed as part of material covered in the CE305: Probability, Statistics, and Data Analysis undergraduate course. Students will have a similar notebook to turn in as part of the weekly assignment, which will be graded for completion and not correctness towards the participation category of their final grade.

Learning Objectives

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?

- Question 2: Suppose you select 50 representative answers to report in the study, what is this subset of answers called?

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

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

- 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?

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:

- Sample Mean is:
- Sample Variance is:
- Sample Standard Deviation:

Events

Basics

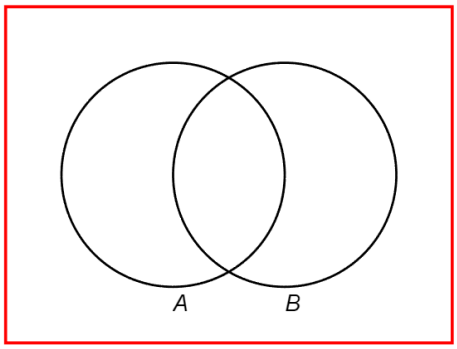
Sample Space: the set of all possible outcomes of a random experiment.

Event:

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

Event: A subset of a sample space.

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 $A_1, A_2, \ldots$ 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:

- Question 1: What is the size of the sample space?

- Question 2: What are the events defined for this sample space?

- Question 3: What is the probability that students will enroll in math OR physics?

- Question 4: What is the probability that students do not enroll in math or physics?

- Question 5: What is the probability that students will only enroll in Math?

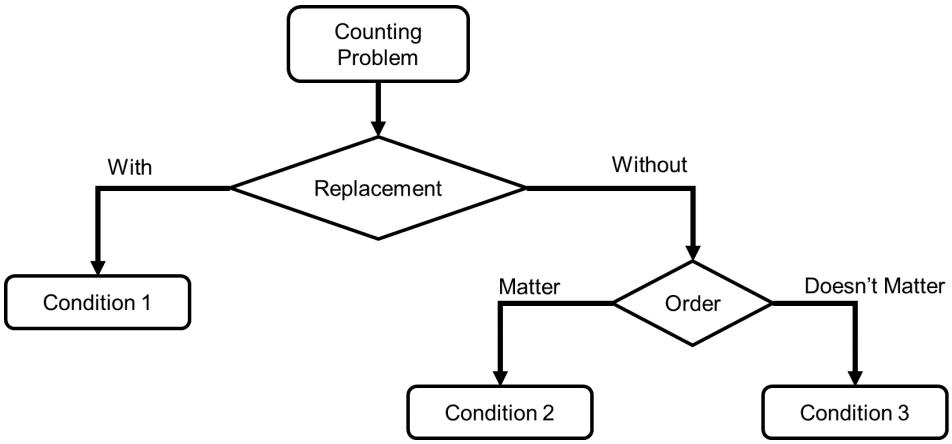
Counting Methods

Basics

Probability: it is the number of times an outcome occurs out of the total possible number of outcomes in a sample space

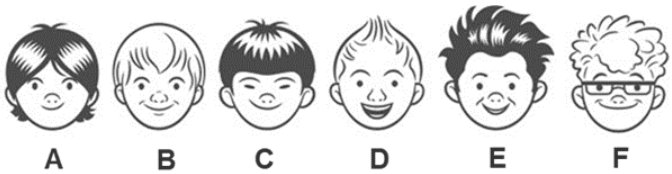
To compute probabilities of events in a sample space, we need efficient ways to count the number of outcomes in the sample space as well as for the event of interest. To this end, we need to identify the nature of the experiment and select the appropriate counting method.

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



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

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?