MATH 108: Elementary Probability and Statistics

Ramapo College of New Jersey

Instructor: Dr. Atul Anurag

Semester: Fall 2025 Homework 1 Due on: September 11, 2025 Date: September 4, 2025

Homework 1: From Section 1.1, complete Problems 1(a, b, c, f, g, h), 2, 11, 13, 15, 17, 19, 21, 23, 25, 35, 37, 58; and from Homework H0, complete part (a).

Quiz 1: The first quiz will be available on Canvas starting Monday and will cover material from Lectures 1 and 2. It must be completed by 8:00 AM on Thursday, September 11.

Review Definitions From Lecture 1

This lecture covers sections 1.3, 1.4, 2.1, & 2.2.

1 Sections 1.3 & 1.4

1.1 Section 1.3: Simple Random Sampling

Random sampling *is the process of using chance to select individuals from a population to be included in the sample.*

Example: Imagine Gallup does a survey about baseball fans by asking people outside Fenway Park in Boston. This wouldn't represent all American adults, because many people there are likely to be baseball fans and not like the average person.

1.2 Sampling Techniques

To avoid bias, several sampling methods have been developed:

- Simple Random Sampling
- Stratified Sampling
- Systematic Sampling

In all these methods, the researcher has no influence over which individuals are selected. We'll begin with the most basic method.

1.3 Simple Random Sampling

A sample of size n from a population of size N is obtained through **simple random sampling** if every possible sample of size n has an equally likely chance of being selected. Such a sample is called a **simple random sample**.

How to Obtain a Simple Random Sample

In small populations, we could write each individual's name on a slip of paper, place the slips into a hat, and draw n names. In practice, populations are usually large, so a more efficient method is used:

1. Assign each individual in the population a unique number from 1 to N.
2. Use a method of random selection (e.g., a table of random numbers or a random number generator) to select n distinct numbers.

Note: To assign numbers, we must first obtain a **frame**—a complete list of all individuals in the population.

Example: Using a Table of Random Numbers

Problem: The accounting firm *Senese and Associates* wants to survey 5 of its 30 clients to assess client satisfaction.

Solution:

1. List all 30 clients (this is the frame) and assign them numbers from `01` to `30`.
2. Use a table of random numbers to select 5 unique numbers.

This method is called **sampling without replacement**, meaning once a client is selected, they are not eligible to be chosen again. In contrast, **sampling with replacement** allows repeated selection. For most surveys, sampling without replacement is preferred to avoid duplication.

Technology

- Using **TI-83/84 Plus**
- Using **TI-84 Plus C**
- Using **Excel**

Steps using Excel

Microsoft Excel can be used to generate random numbers for simple random sampling using the **Data Analysis ToolPak**. Follow the steps below:

1. First, ensure the **Data Analysis ToolPak** is activated:
   - Go to `File` → `Options`.
   - Select `Add-Ins`.

- Under `Manage:`, choose `Excel Add-ins` and click `Go...`.
- Check the box labeled `Analysis ToolPak` and click `OK`.

2. On the ribbon, select `Data`, then click on `Data Analysis`.
3. In the dialog box, highlight `Random Number Generation` and click `OK`.
4. Fill out the Random Number Generation window with appropriate values. For example, to simulate Example 2 (selecting a random sample of 5 from 30 clients), do the following:
   - **Number of Variables:** 1
   - **Number of Random Numbers (rows):** 10
   - **Distribution:** Discrete Uniform
   - **Value Between:** 1 and 31 (inclusive)
   - **Output Range:** Enter the cell where you'd like the numbers to appear (e.g., `A1`)
5. Click `OK`. Excel will generate the random numbers in the specified column.
6. From the list of 10 numbers, select the first 5 **unique values between 1 and 30**. Discard any duplicates and any value equal to 31.

Note: More than 5 numbers are generated to ensure that, after removing duplicates or out-of-range values, 5 valid numbers remain.

2 Other Effective Sampling Methods

Goals of Sampling

The goal of sampling is to gather as much meaningful information about a population as possible while keeping costs low. Here, cost includes not only money but also time, effort, and other resources.

While simple random sampling is a basic and important method, other sampling techniques can sometimes be more effective. Below are three alternative methods that may work better in specific situations:

- **Stratified Sampling:** The population is divided into groups (called strata) based on a characteristic (e.g., age, gender), and a random sample is taken from each group. This ensures representation from each subgroup.
- **Systematic Sampling:** A starting point is chosen at random, and then every kth member of the population is selected. This method is simple and often easier to implement than a random sample.

2.1 Obtain a Stratified Sample

Under specific conditions, **stratified sampling** can be more effective than simple random sampling. It is particularly useful when the population contains distinct subgroups that should be represented in the sample.

A **stratified sample** is obtained by dividing the population into nonoverlapping subgroups called **strata**, and then taking a simple random sample from each stratum. The individuals within each stratum should be **homogeneous** (i.e., similar) with respect to some characteristic.

Example: The owner of a private food store is concerned about employee morale. She decides to survey the managers and hourly employees to see if she can learn about work environment and job satisfaction. There are 8 managers and 21 hourly employees. How should she go about selecting 6 people to survey?

To select 6 people for the survey fairly and in a way that represents both managers and hourly employees, the store owner should use a ***stratified random sample****. This method ensures that both groups are proportionally represented in the survey.*

Step-by-step Approach

Identify the strata (groups):

- Managers: 8 people
- Hourly employees: 21 people
- Total employees: $8 + 21 = 29$

More Examples

Suppose a university wants to survey student opinions on campus services. Rather than sampling students at random (which may result in overrepresentation of certain class years), the administration divides students into four strata based on class year. Then, a simple random sample is taken from each class year to ensure balanced representation.

This example describes a stratified random sampling method.

2.2 Obtain a Systematic Sample

Unlike simple random sampling and stratified sampling, which require a **frame** (a list of all individuals in the population), **systematic sampling** does not require such a list. This makes it a practical choice when a complete population list is unavailable.

A **systematic sample** is obtained by selecting every k^{th} individual from the population. The starting point is determined by randomly selecting a number between 1 and k. To obtain a systematic sample, follow these steps:

1. Decide on the sampling interval k (for example, $k = 8$).
2. Randomly select a starting number between 1 and k.
3. Survey the individual corresponding to that starting number.
4. Then, survey every k^{th} individual thereafter (i.e., the individual at position $start + k$, $start + 2k$, and so on) until the desired sample size is reached.

Example: Suppose you want to survey every 8th individual in a population. You randomly pick the number 5 as the starting point. The individuals surveyed will be the 5th, 13th, 21st, 29th, and so forth.

Example: Obtaining a Systematic Sample Without a Frame

Problem: The manager of Kroger Food Stores wants to measure the satisfaction of the store's customers. Design a sampling technique to obtain a sample of 40 customers.

Approach: Since it is difficult (if not impossible) to obtain a frame listing all Kroger customers, systematic sampling is an appropriate choice. The manager can survey every k^{th} customer exiting the store.

Solution: The manager decides to survey every 7th customer. He randomly selects a starting number between 1 and 7, say 5. He then surveys the 5th customer leaving the store and every 7th customer thereafter until he reaches a sample size of 40 customers. The customers surveyed will be the 5th, 12th, 19th, 26th, ..., 278th customers.

This method provides a practical and efficient way to collect data without requiring a complete list of all customers.

Choosing the Sampling Interval k in Systematic Sampling

A critical step in systematic sampling is selecting the value of k, the sampling interval. This choice affects whether the sample is representative and whether the desired sample size can be achieved.

When the Population Size N Is Unknown If the size of the population is unknown, there is no precise mathematical way to determine k. Instead, the value of k must strike a balance:

- It must be **small enough** to ensure that the desired sample size can be reached.
- It must be **large enough** to produce a sample that adequately represents the entire population.

Example Revisiting the Kroger example from before:

- If k is too large (e.g., 30), surveying every 30th customer starting at the 5th would require that at least 1175 customers visit the store for a sample size of 40 to be obtained. If fewer customers visit, the sample size will be insufficient.
- If k is too small (e.g., 4), surveying the 5th, 9th, ..., 161st customer might miss some customer segments—such as evening shoppers—resulting in a non-representative sample.

In such cases, an estimate of the population size helps in choosing an appropriate k.

When the Population Size N Is Known If the population size N is known, selecting k is straightforward:

$$k = \left\lfloor \frac{N}{n} \right\rfloor,$$

where n is the desired sample size and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

For example, if $N = 20,325$ and $n = 100$, then

$$k = \left\lfloor \frac{20,325}{100} \right\rfloor = 203.$$

A random starting number p is selected between 1 and k, and the sample consists of individuals numbered:

$$p, \quad p+k, \quad p+2k, \quad \ldots, \quad p+(n-1)k.$$

If $p = 90$, the sampled individuals would be the 90th, 293rd, 496th, ..., 20,187th individuals.

Summary: Steps in Systematic Sampling

1. If possible, approximate the population size N.
2. Determine the desired sample size n.
3. Compute the sampling interval $k = \left\lfloor \frac{N}{n} \right\rfloor$.
4. Randomly select a number p between 1 and k.
5. Select individuals numbered $p, p+k, p+2k, \ldots, p+(n-1)k$.

Note: Because we are surveying 40 customers in the Kroger example, the first individual surveyed is the 5th, the second is the $5+7=12$th, the third is the $5+2\times 7=19$th, and so on, until the 40th, which is $5+39\times 7=278$th customer.

End of Lecture #2