

1.2 Sampling

- **Census and Parameters -vs- Samples and Statistics**

- When you gather data from every member of a population it is called a **census** and the results are called **parameters**.
- When you gather data from a subgroup of a population it is called a **sample** and the results are called **statistics**.

- **Random and Simple Random Samples**

- In a **random sample** every member of the population has an equal chance of being selected.
- A **simple random sample** is a random sample where every sample of the same size has the same chance of being selected. There can be no sub-grouping of the population.
- A simple random sample is always random but the reverse is not necessarily true.

- **Examples:**

Classify each sampling method as simple random, random but not simple random, or neither.

1. In a class with 200 males and 300 females, I want to select 50 students for a survey.

(a) I randomly select 20 males and 30 females for the survey.

Answer: This is random because all students have a $1/10$ chance of being picked, but it is not simple random because I can't have a sample of say 25 males and 25 females.

(b) I put all 500 students in a list and randomly select 50 students.

Answer: This is simple random sample because everyone has an equal ($1/10$) chance of being picked **and** no sample of 50 has a better chance of being picked.

(c) I randomly select 25 males and 25 females.

Answer: Neither. Males have a $25/200$ chance of being picked and females have a $25/300$ chance of being picked. Not all students have the same probability of being picked so it is not even a random sample.

2. **Your Turn:** In my class I have 30 students, there are 5 rows of 6 students each. I want to select 12 students for a survey. Classify each sampling method as simple random, random but not simple random, or neither.

(a) I put all 30 students names in a basket and randomly select 12 students.

(b) I randomly select two of the five rows and choose all students in each of these rows.

(c) I randomly select one odd row and one even row and take all students in the chosen rows.

- **Sampling Strategies**

- In a **systematic sample**, every n^{th} member of the population is selected.
- In a **convenience sample**, the most convenient subgroup is selected.
- A **stratified sample** is one in which the population is divided into two or more sub-groups, called strata, that share similar characteristics. We then draw a random sample from each strata. This is good if you want to make sure that some members from all strata are present in the sample. Note: A stratified sample may be random but it won't be simple random.
- In **cluster sampling**, we divide the population into groups (or clusters), then randomly select some of those clusters. Once a cluster is selected all the members of that cluster are included in the sample. This method is often used for convenience purposes. Note: Again, a cluster sample may be random but it won't be simple random.

- **Examples:** Classify each sampling method as **systematic**, **convenience**, **stratified**, **cluster**, or **none of these**. Does the method constitute a random sample? If it is random, is it simple random?

1. You're considering a lunch-delivery business and want to gather lunch-break data on a sample of employees from your target population of 20 local businesses.

(a) You randomly select 3 of the businesses and interview all the employees from those businesses.

Answer: This is a cluster sample (the employees are clustered by the business employing them). It is random because all members of your target population have a $3/20$ chance of being selected. It is not simple random because you have grouped your subjects prior to sampling.

(b) You get a sample of 100 by randomly selecting 50 employees from labor and 50 employees from management.

Answer: This is a stratified sample (the strata are labor and management). It is unlikely to be random unless there are an equal number of labor and management employees in your target population. Either way, it is not simple random because not all samples of 100 have the same probability of being selected.

2. **Your Turn:** Suppose you want to gather household income information from a sample of 10 houses on North Street. The house numbers start at 1 and end at 100 without any missing house numbers.

(a) You randomly select 5 even numbered houses and 5 odd numbered houses.

(b) You take every 10th house starting at number 7.

(c) You group the houses as #'s 1 - 10, 11 - 20, ..., 91 - 100. You then randomly select one of these groups to be in the sample.

(d) You randomly select 10 numbers between 1 and 100 and select those 10 houses.

(e) You include the first 10 houses where someone answers the door.