

Chapter 2: Samples, Good and Bad

Overview: In this section, we begin to explore sampling as a method used to estimate features of a population. We will talk about a few different kinds of samples, some which are biased and some that seek to eliminate any bias.

Motivating Example: Suppose we want to determine the prevalence of a certain disease (I think we can all think of one that is of concern these days) in Canada. In order to actually determine the true value of this proportion, we would need to test every single person in the population for the disease with a test that is 100% accurate. That would be incredibly expensive and essentially impossible to execute. Instead, we could take a **sample** of Canadians and test them for the disease. How do we choose who is in the sample? What if someone selected for the sample refuses to participate in the study?

Types of Samples: We now define 4 different kinds of samples:

- A convenience sample \Rightarrow select whichever individuals that are easiest to reach.

e.g. conduct a survey by standing in a fixed location and ask passers by.

problem: sample may not represent the population.

- A voluntary response sample chooses itself by responding to a general appeal.
Same problem as above.

e.g. voting, polling

- A simple random sample (or SRS) n individuals

\Leftrightarrow those n individuals are chosen in the population in such a way that they are equally likely to be selected.

e.g. population of 14 students in this class.
want a random sample of 3.

1. Assign each individual with a number 1 to 14.
2. Pick three numbers "randomly"

- A stratified random sample

1. Divide the population into subgroups (strata)
2. In each stratum, we do a simple random sample.

Stay tuned

Practice Question: Suppose I want to determine the proportion of current UVic students who plan on taking Stat 353 at some point in their degree. I want to collect a sample from UVic's population and I have a few ideas on how to take this sample.

In each scenario described below, determine what kind of sampling is being described.

1. I get a list of all of the students numbers of current UVic students (as well as the corresponding email addresses) and randomly select 200 of them to contact and ask this question.

(A) Convenience (B) Voluntary Resp. (C) SRS (D) Stratified Random

2. I ask everyone in Stat 123 to answer this question on their first homework assignment and use the entire class as my sample.

(A) Convenience (B) Voluntary Resp. (C) SRS (D) Stratified Random

3. I post signs around UVic asking students to contact me with their answer to the question.

(A) Convenience (B) Voluntary Resp. (C) SRS (D) Stratified Random

strata
↓

4. I randomly sample 5% of the students from each department at UVic.

(A) Convenience (B) Voluntary Resp. (C) SRS (D) Stratified Random

Certain types of sampling can produce something called bias. We say that a sample is biased if: it systematically favours certain outcomes.

A biased sample is one that is NOT representative of the population.
convenience, voluntary (biased)

Almost always, convenience sampling and voluntary response sampling produce a biased sample. It is best to do some type of random sampling.

Question: When should you use a Simple Random Sample vs a Stratified Random Sample?

Answer: Stratified sampling should be used when you believe that each stratum will be different from the overall population.

Simple random sampling should be used if you think that all subgroups are about the same.

Note: There are other types of random sampling that we did not define here which are explored further in Stat 354. In this course, we will focus on simple random sample only.