

Sampling Principles and Strategies

Colby Community College

Definition

A **population** is the complete collection of all measurements or data that are being considered.

Example 1

A researcher asks:

What is the average mercury content in swordfish in the Atlantic Ocean?

What is the target population?

The target population is all the swordfish in the Atlantic ocean.

Definition

A **census** is the collection of data from every member of a population.

Note

It is often prohibitively expensive to collect data from every member of a population.

Definition

A **sample** is a subset of members selected from a population.

Example 2

In a journal article it was stated that there are 38 million carbon monoxide detectors installed in the United States. When 30 of them were randomly selected and tested, it was found that 12 of them failed an alarm in hazardous carbon monoxide conditions.

What is the population?

All 38 million carbon monoxide detectors in the United states.

What is the sample?

The 30 carbon monoxide detectors that were selected and tested.

Ryan, T. J., & Arnold, K. J. (2011). Residential carbon monoxide detector failure rates in the United States. American journal of public health, 101(10), e15–e17. <https://doi.org/10.2105/AJPH.2011.300274>

Definition

Anecdotal evidence is a factual claim relying only on personal observation, collected in a casual or non-systematic manner.

Example 3

A man on the news got mercury poisoning from eating swordfish, so the average mercury concentration in a swordfish must be dangerously high.

Example 4

My friend's dad had a heart attack and died after they gave him a new heart disease drug, so the drug must not work.

Note

Such evidence may be true and verifiable, but it may only represent extraordinary cases.

Example 5

Suppose we as a student who happens to be majoring in nutrition to select several graduates for a study.

What kind of students do you think she might collect?

Do you think her sample would be representative of all students?

Definition

When data is gathered in a way that is not truly random, the sample is considered **biased**.

Note

Any time samples are selected by hand, there is a risk of picking a biased sample, even if the bias is not intentional.

Note

Sample selection is often done by computer.

Definition

A **voluntary response sample** (or **self-selected sample**) is one which the respondents themselves decide whether to be included.

Example 6

Internet polls, such as on Twitter, where people decided whether or not to participate.

Example 7

Call-in polls, where people are asked to call a special number to register an opinion.

Definition

A **loaded question** is one that is intentionally worded to elicit a desired response.

Example 8

When asked

“Should the President have the line item veto to eliminate waste?”

97% of respondents said yes.

Example 9

When asked

“Should the President have the line item veto?”

57% of respondents said yes.

Definition

A **non-response** occurs when someone refuses to respond or is not available.

Example 10

Many telemarketers disguise their sales pitch as an opinion poll, causing the non-response problems to increase in recent decades.

Example 11

One major hurdle for the U.S. Census is getting marginalized groups to respond. Since the number of congressional districts are determined by the results of the census, marginalized groups are often under-represented in government.

Definition

A **convenience sample** is one where individuals who are easily accessible are more likely to be in the sample.

Example 12

If a polling company based in New York City only stopped people walking outside their office, the sample would not be representative of the entire city.

Example 13

As of 2021, BIC, which is headquartered in France, has 25 factories around the world.

If an executive wanted to study how efficient the company manufactures pens, but only collected data in the French factories, the sample would not be representative of all pens the company makes.

Definition

A **prospective study** identifies individuals and collects information as events unfold.

Example 14

Harvard, in 1938, began studying a group of 268 sophomores, and soon an additional study added 456 inner-city Bostonians. To this day, they are surveyed every two years and physical examinations given every five.

Definition

A **retrospective study** collects data after events have taken place.

Example 15

Hospitals will at times perform chart reviews, where they gather the records of all patients that underwent a specific procedure. They will then review the documents and make recommendations.

Definition

A **confounding variable**, also called a **lurking variable**, is a variable that is correlated with both the explanatory and response variables.

Note

Ideally, researchers would find all confounding variables, but there is no guarantee that all confounding variables can be examined or measured.

Example 16

Suppose a study tracked sunscreen use and skin cancer, and it was found that the more sunscreen someone used, the more likely the person was to have skin cancer.

Does this mean sunscreen causes cancer?

This study didn't track sun exposure, which is a confounding variable. If someone is in the sun all day, they are more likely to both use sunscreen and get skin cancer.

Definition

A **simple random sample** is selected in such a way that every individual the same chance of being chosen.

Note

While random sampling may seem easy, it actually takes a great deal of planning to implement correctly.

Example 17

The NFL is made up of 32 teams and hundreds of players. If we wanted to take a simple random sample, we could write the name of every player on a piece of paper and draw 128 names out of a hat.

Example 18

Suppose a polling company maintains a list of all phone numbers active in the country. When they want to conduct a survey, a computer will randomly pick phone numbers from this list to call.

Definition

In **stratified sampling** the population is divided into groups called **strata**. Then a simple random sample is selected from each strata.

Note

The strata are usually chosen so that similar cases are grouped.

Example 19

Each team in the NFL has a different amount of money to spend on salaries. Instead of picking 128 players completely at random, we could pick 4 players from each of the 32 teams.

Example 20

To ensure that a sample is half men and half women, researchers can group all women in one strata and all men in another. They then randomly pick 50 from each strata, giving a sample with 100 people.

Definition

In **cluster sampling** we break up the population into groups, called **clusters**. Then we randomly select a fixed number of clusters and include all individuals in the selected cluster in our sample.

Example 21

To sample a city, each neighborhood could represent a cluster and then use a simple random sample to pick clusters.

Note

For cluster sampling to work well, all the clusters need to be similar. If there are significant differences between the clusters, the sample may not be representative of the population.

Example 22

Suppose an eraser factory produced 1000 lots a day. To perform quality control, the company randomly selects 5 lots and tests all erasers in the chosen lots.

Definition

A **multistage sample** is like a cluster sample, but rather than keeping all individuals in each sample, we collect a random sample in each cluster.

Example 23

Suppose we want estimate the malaria rate in rural Indonesia. We learn there are 30 villages in that part of the country.

We decide to randomly select half of the villages. Then from each village we randomly select 10 people.

This gives us a sample of 150 people, but we only need to pay to travel to half the villages in the area.

Note

We will primarily be considering simple random samples in the course, since advanced methods needed to analyze the more complicated sampling methods.