Sampling

Time	8 minutes
Description	Sampling is how we study a group without surveying everyone. In this module, you'll learn how sample size and sampling methods shape the accuracy and fairness of your results. Get it right, and your data speaks for the population—get it wrong, and your conclusions could be misleading.
# Module	1

The groups you choose to collect data from can make or break your results.

This module explores two key concepts that shape the accuracy, fairness, and reliability of your data:

- Sampling Size
- Sampling Methods

By understanding these concepts, you'll be able to collect **representative** data and make **smarter decisions**.

Summary

- **Sampling methods** (probability and non-probability) help you select who's in your study
- The **sample size** influences how much your results reflect the bigger population
- Choosing the **right approach** helps reduce bias and draw stronger conclusions

Sampling Size

> 💡 **Sample size**
>
> How many people or things you include in your analysis—and how that number affects your ability to see patterns and make generalizations.

Why Sample Size Matters

- A sample of 1 person won't tell you much about your entire city.
- A sample of 10,000 people is more likely to reflect patterns and differences, especially when trying to generalize results.
- Not every project needs 10,000!

Goal	Suggested Sample Size
Local neighborhood survey	20-100
Statewide poll	1,000+
National public opinion	2,000-3,000
Classroom project	15-30

> 💡 **Law of Large Numbers**
>
> The greater the sample, the more accurate your results get.

- Flip a coin 3 times → you might get 3 heads.
- Flip it 1,000 times → you'll likely get closer to 50/50 heads and tails.

Sample Size in Context

Some situations don't need a huge sample:

- 20 neighbors might be enough to understand recycling habits in your neighborhood.
- But you'd need thousands to generalize for your entire state.

Statistical Significance

> 💡 **Statistical Significance**
>
> This tells you whether your result is likely real or just due to random chance.

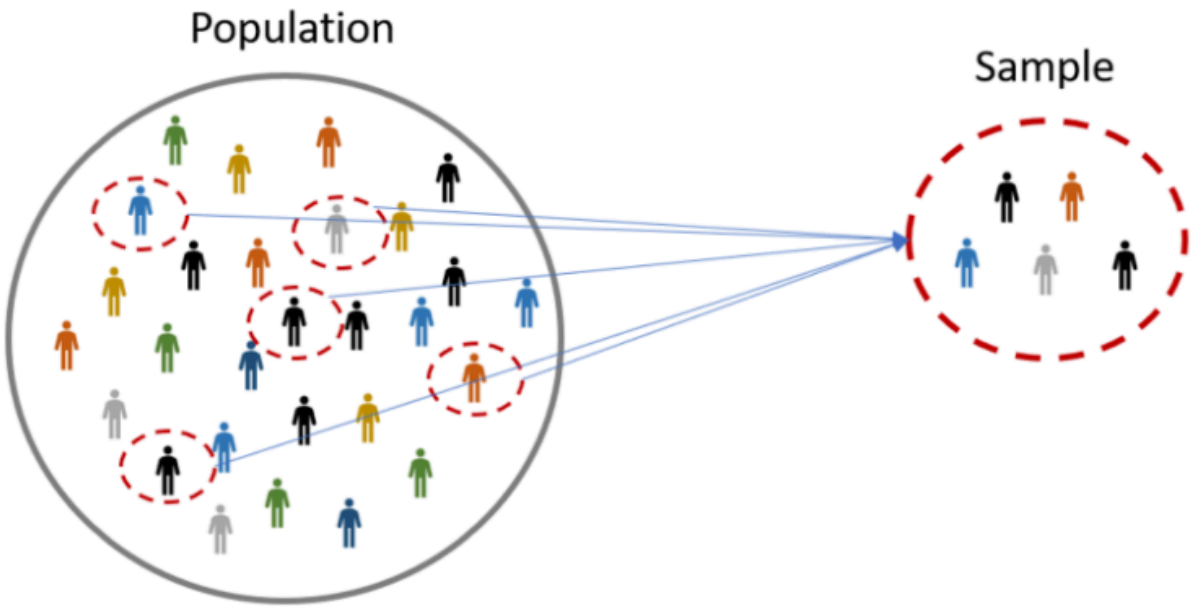
- A larger sample size helps detect small, but meaningful differences
- But what "meaningful" looks like depends on the context

Example

- A 10% difference in favorite ice cream flavor = interesting
- A 10% difference in cancer treatment outcomes = critical

💡 **Magnitude of Difference**

How big the difference is between two groups. Bigger differences are easier to detect.



via: https://www.omniconvert.com/what-is/sample-size/

While having larger sample sizes may generally be better, you also have to consider who is going to be sampled. Thus you should also consider **sampling** methods.

Sampling Methods

💡 **Sampling Methods**

The process of choosing who you include in your data.

You usually can’t study everyone. So instead, you pick a sample — and how you pick it matters. Poor methods can lead to biased results.

Probability Sampling

Methods where everyone has a known, random chance of being picked. Reduces bias and increases reliability.

Method	How it works	Real-Life Example
Simple Random	Everyone has an equal chance to be picked.	Drawing names out of a hat
Systematic Random	Pick every nth person from a list	Every 5th person in line
Stratified Random	Split into groups, sample from each group	Randomly pick from each age group
Cluster Random	Randomly choose whole groups, then study everyone in them	Survey all patients from 3 randomly chosen clinics

via https://www.scribbr.com/methodology/sampling-methods/

Non-Probability Sampling:

A method where not everyone has a chance to be picked. Easier to do, but more risk of bias.

Method	How it Works	Risk Example	Potential Bias
Convenience	Researcher chooses a sample that is available in a non-random way	Researcher asks people on a college campus at 1:30 pm questions as they walk by.	More likely to get college students, as time of day is during the work day and on college campus.

Voluntary Response	Researcher asks people to join the sample and the people decide if they want to or not.	A restaurant asks you to answer questions about service at the end of the meal.	Open to bias because people that want to respond will likely have stronger opinions than everyone else.

Be cautious! Non-probability samples may leave out important voices or over-represent strong opinions.

Takeaways

Sampling is how we make smart guesses about a population by collecting data from a smaller group.

To get the most accurate picture:

- Choose a **sample size** that fits your goals
- Use the **best sampling method** for your goals
- Always consider **who's missing** from data

> 💡 A **representative sample** is one that reflects the larger population accurately.

Additional Resources:

- Sample Size Explanation
- Sampling Methods