Quantitative Research Methods SS24

Sampling

Constantin Lackinger & Philipp Pilatowicz

STAND: MARCH 2024

What is sampling?

Sampling is the process of selecting **units** (e.g., people, organizations) from a **population** of interest so that by studying the sample one can **generalize results back to the population** from which they were chosen.

Sampling Terminology

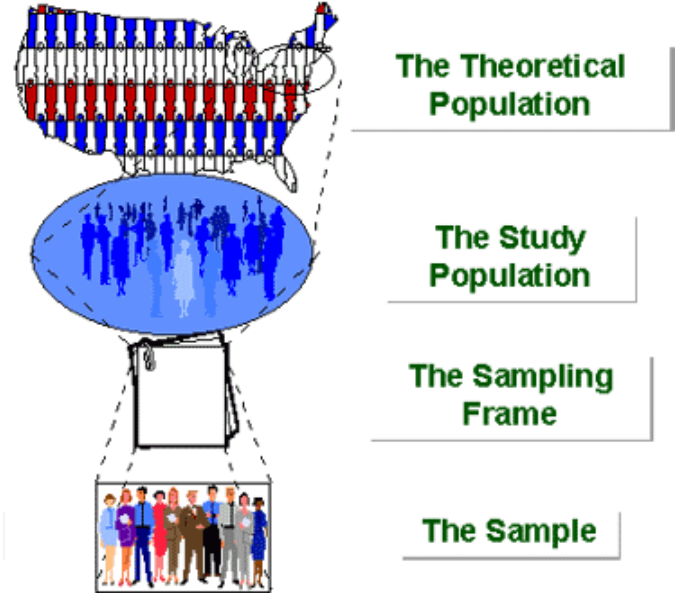
Population: the entire group you're interested in making conclusions about.

- Theoretical population
- Accessible population

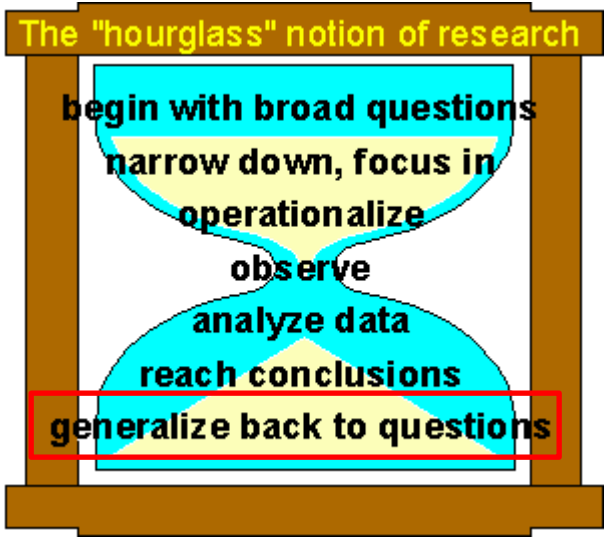
Census: official survey involving the entire population of interest.

Sampling Frame: a list of every unit in your population from which the sample is drawn (e.g., a phonebook list) or a procedure that approximates such list giving you access to the units in your population

Sample: the group you're selecting to studying (=/= the group you actually study)



Why is sampling of concern?



Individuals are often studied, but usually there is interest in generalizing and drawing conclusions beyond just the individuals studied.

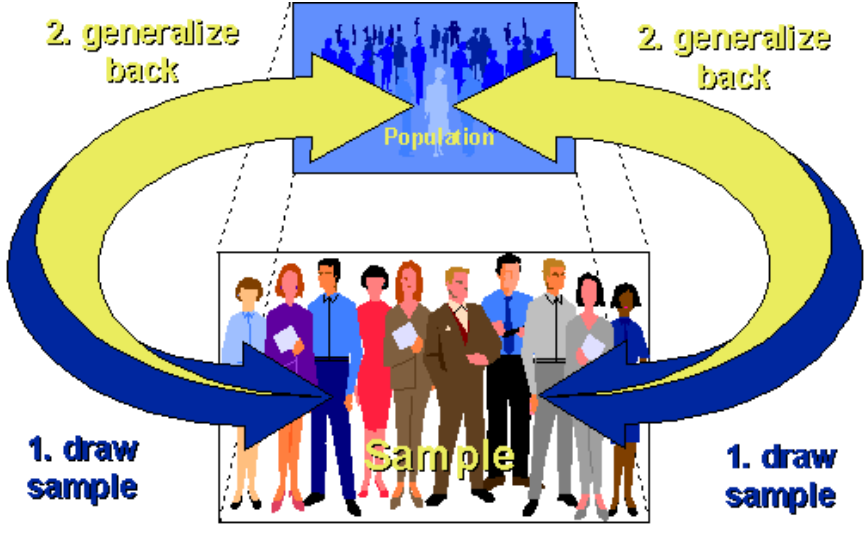
- But how can you generalize conclusions and fulfill the external validity criteria?
- To what or whom to you want to generalize to?
- Does the generalization of the results of the study hold for other settings, people, places and times?
- Can we even generalize with certainty?

Approaches to generalization & external validity

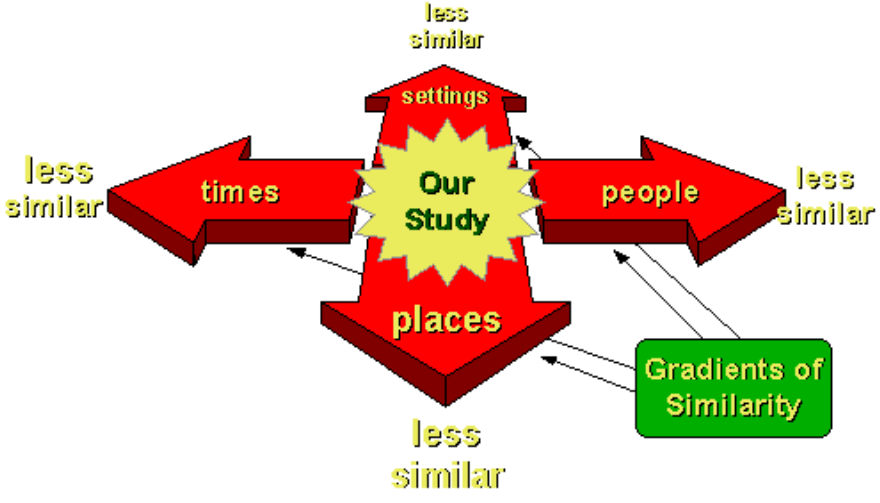
Sampling model: Identifying population of interest and drawing a sample that reflects this population.

Issues:

- Potentially unknown population
- Drawing representative sample issues
- Time generalizability (e.g., predict future)



Approaches to generalization & external validity



Proximal similarity model: thinking about different contexts and identify which contexts are more like your study and which are less so (→gradient of similarity & extend of generalizability)

Issue:

- No certainty of generalizability (only proximally similarity)

Threats to generalization & external validity

Explanation of potential inaccuracies in generalizations:

- Three (or four) main threats: people, places, and times (and settings).
- Critiques may argue that results are influenced by unique characteristics of the selected study's participants, location, or timing, (or setting).

Improving External Validity (=generalizability of your study conclusion to other persons in other places in other times):

- Utilize random selection for sampling over nonrandom sampling to enhance representativeness.
- Minimize dropout rates to maintain sample integrity.
- Invest time in clarifying and justifying your context and how this context is similar/different to other contexts (e.g., concept mapping to visualize contextual similarities, provide data, …)
- Strengthen external validity through replication across diverse contexts, demonstrating consistency and robustness of findings.

Issues on the way introducing bias

Issues identifying theoretical population:

Who do you want to generalize to?

Issues accessing identified population:

What population can you get access to?

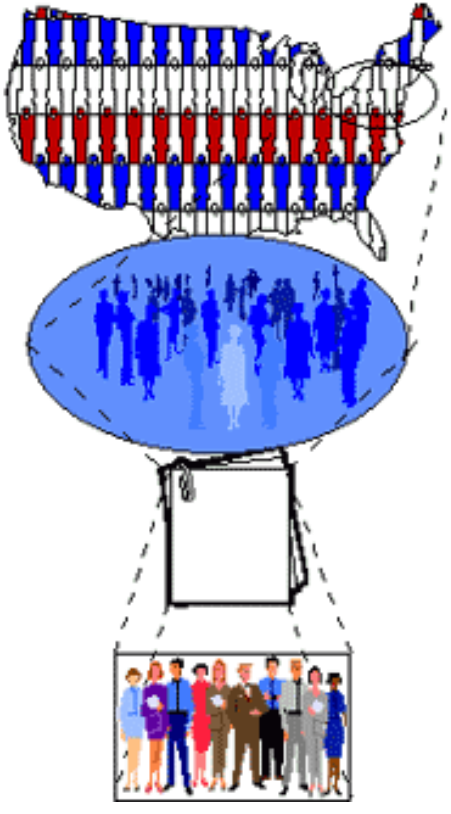
Issues regarding accuracy of sample frame:

How can you get access to them?

Issues drawing accurate and correct sample:

Who is in your study?

Issues of units participating accurately and correctly:



The Theoretical Population

The Study Population

The Sampling Frame

The Sample

Sampling Distribution

WU WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Responses are collected, you can now use a statistic (mean, median, mode) to analyse. **But how can you estimate to the generalized population?**

Sampling Distribution

Sampling Distribution:

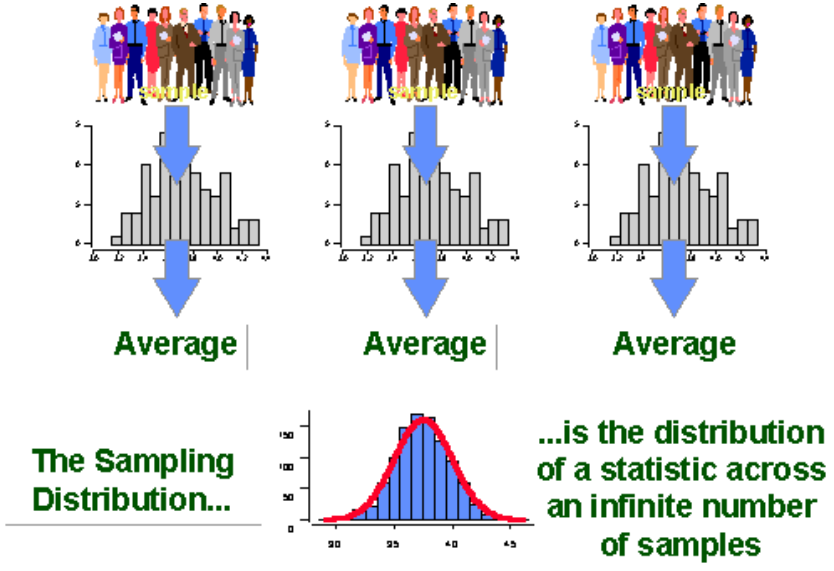
The distribution of a statistic across an infinite number of samples.

In easier words:

- Suppose that you draw a random sample from a population and calculate a statistic for the sample, such as the mean.
- Now you draw another random sample of the same size, and again calculate the mean.
- You repeat this process many times, and end up with a large number of means, one for each sample.

The distribution of the sample means is an example of a **sampling distribution.**

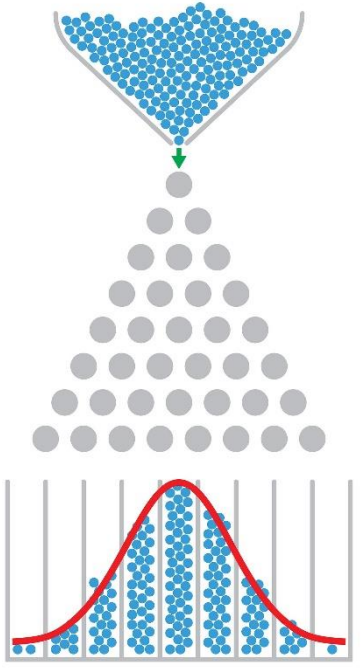
→ This graph will always look like a bell curve



Don‘t believe it?

The Galton Board and the central limit theorem:

- The central limit theorem relies on the concept of a sampling distribution.
- The central limit theorem says that the sampling distribution of the mean will always be **a bell curve**, as long as the sample size is large enough.
- The Galton Board easily demonstrates this important concept.



Standard deviation & standard error

Standard deviation: the spread of the scores around the average in a single sample (variability within your sample).
Standard error: is the spread of the averages around the average of averages in a sampling distribution (variability across multiple samples).

- Indicates how well your sample data represents the whole population.
- The larger the sample, the smaller the error as you approach the population itself.
- **Example standard error of the mean:** Your infinite sample means may each deviate from the actual mean of a population; this deviation is the standard error of the mean.

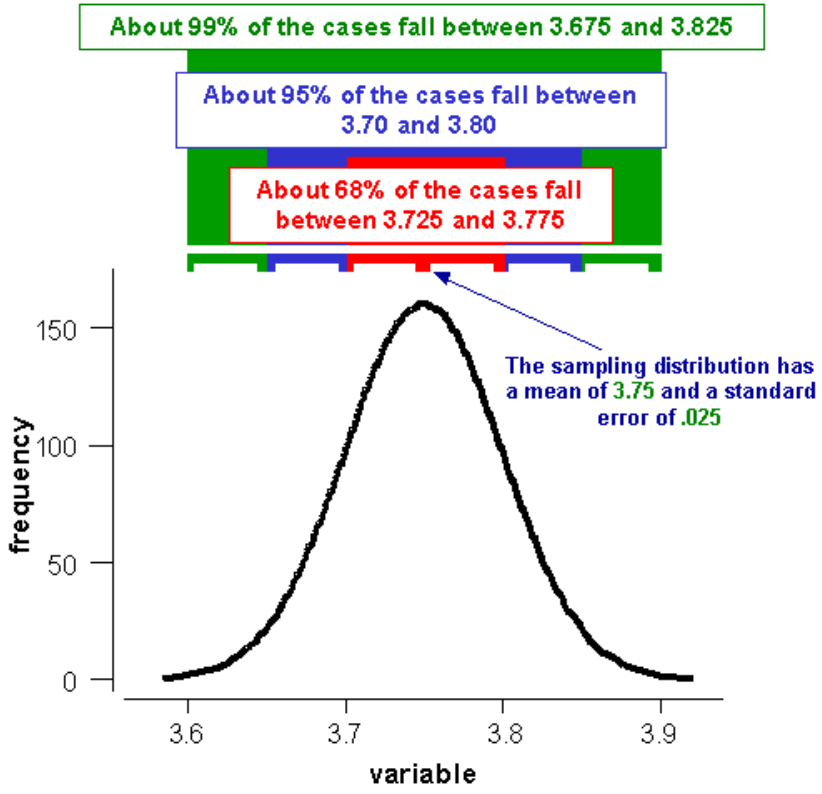
Sampling error: the standard error in sampling.

Confidence intervals

- When you make an estimate in statistics there is always uncertainty around that estimate because the number is based on a sample of the population you are studying.
- However, we might **estimate the range** within which the **true population parameter is likely to fall.**

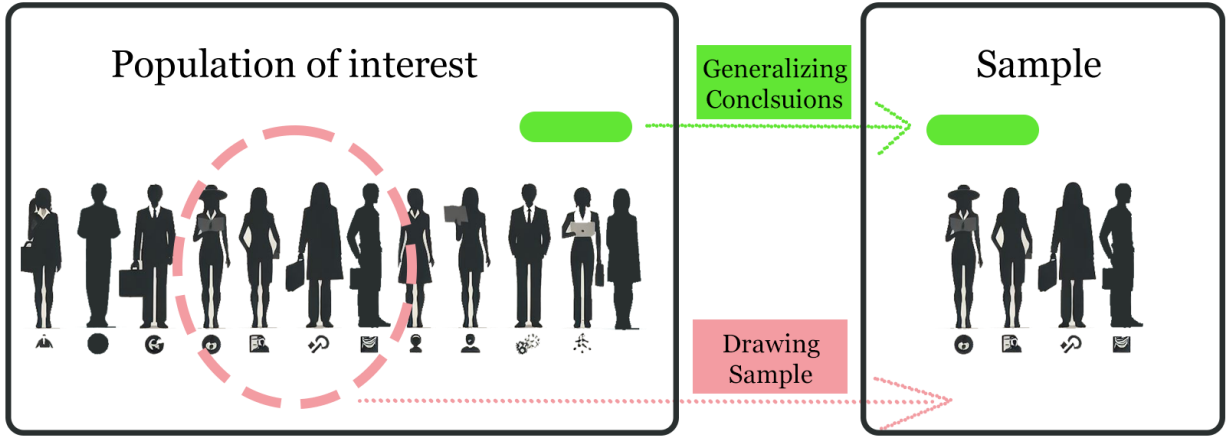
Confidence interval is the range of values that you expect your estimate to fall between a certain percentage of the time if you re-sample the population in the same way. This range is typically employed to predict within which 68%, 95%, and 99% of cases would be expected to fall.

This enables generalizability within a range (with a 68, 95 or 99% confidence)



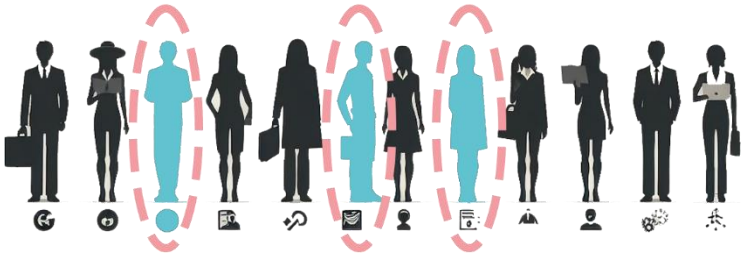
Sampling Methods
Probability Sampling
(random selection, equal probabilities)

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS



Simple Random Sampling
Keep it simple.

Conducting a sample from a population in which each potential sample has an identical chance of being chosen.



All units have the same chance of being selected (e.g. 3/12 = 25%)

- Excel function *=RAND()*
- Statistical tools
- Mechanical tools (ball machine)

Advantages

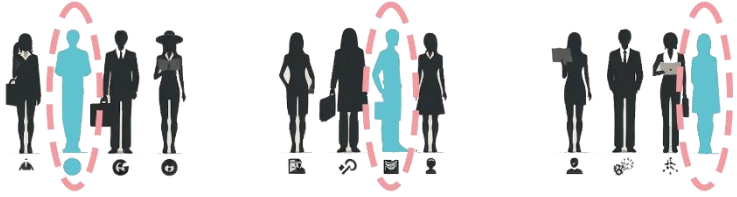
- easy to accomplish and explain to others, universally used

Disadvantages

- Requires full sample list
- Can be inefficient/time-consuming (depending on the tool)
- Sample Selection Bias: Weak representation of sub-groups

Stratified Random Sampling
Capturing minority groups.

Dawing a sample out of several homogenous sub-groups of a population and then taking simple random sample out of sub-groups



Decide between:

- **Proportional stratified** = same sampling fractions
- **Disproportional stratified** = different sampling fractions (for oversampling certain groups or minorities)

Advantages

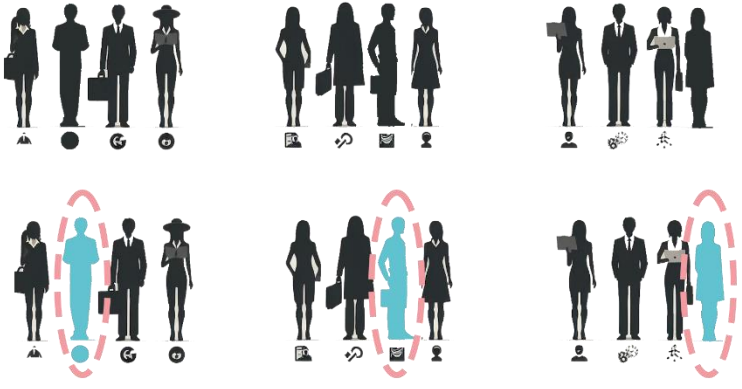
- Allows oversampling of minority groups
- Enhances representation of sub-groups

Disadvantages

- Requires full sample list

Systematic Random Sampling
Tackling large sampling frames.

Drawing a sample based on a rule to select each x. element of randomly ordered sampling frame starting from a random position in the list



Starting at random position and chose every e.g. 5th unit out of randomly ordered sampling frame

Advantages

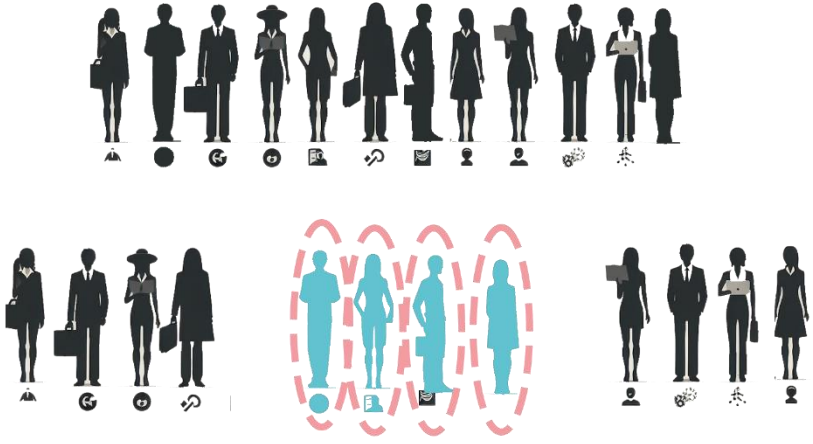
- No full sample list needed and counting through it to find randomly selected ones

Disadvantages

- Risk of bias if order is non-random

Cluster (Area) Random Sampling
Dealing with geographically dispersed populations

*Dividing the population into (geographical) clusters, randomly selecting a few clusters and then using **each** element in these clusters for the sample*



Combined with other methods (such as simple random sampling)

Advantages

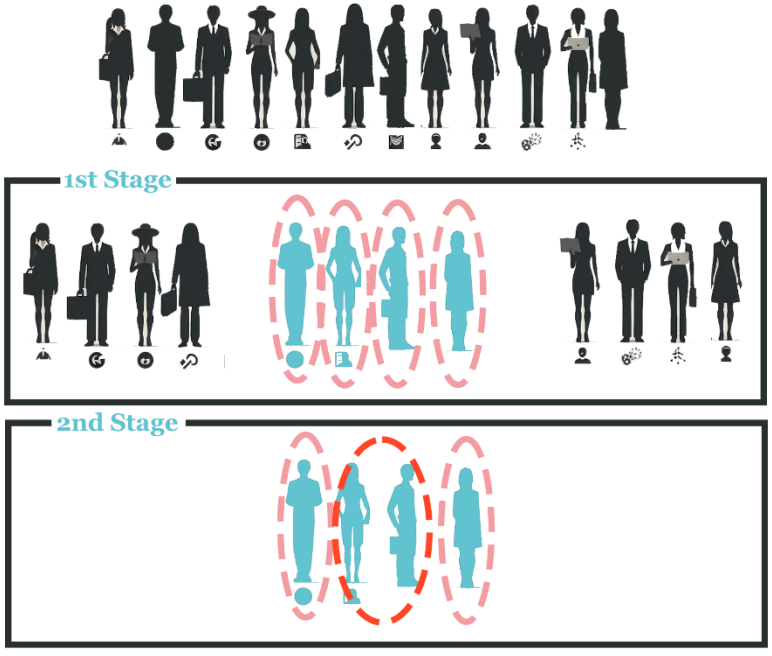
- More efficient when sampling across (geographically) high dispersed area/population

Disadvantages

- Usually not used alone, only in combination with other methods

Multistage Random Sampling
Complex, but sophisticated.

Combining several sampling techniques for different stages of sampling (e.g. cluster sampling followed by simple random sampling)



Advantages

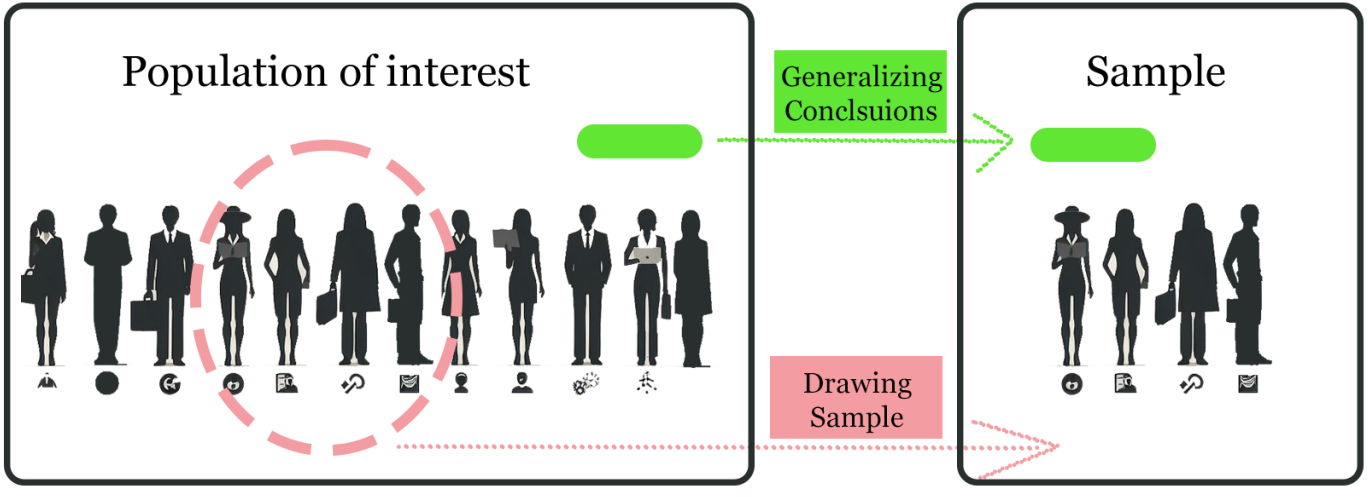
- Highly sophisticated and precise method
- Efficiency for complex population to sample processes

Disadvantages

- Can get complex with multiple stages
- Difficult to explain to audience

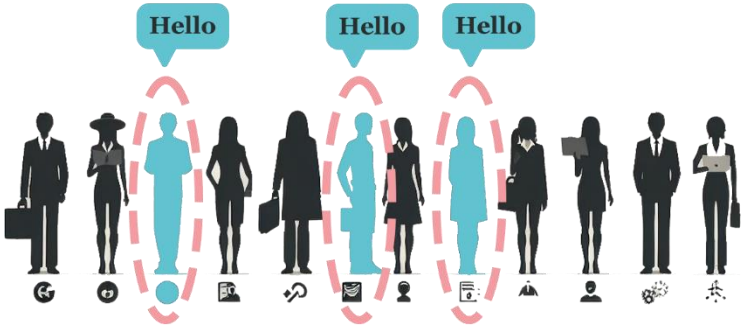
Sampling Methods
Non-Probability Sampling
(non-random selection)

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS



Convenience Sampling
Taking what is accessible.

Drawing a sample based on units you randomly approach, are available or willing to participate in the study (e.g. person-on-the-street interviews)



Method based on convenience, available ressources, accessible units rather than reaching high representativity

Advantages

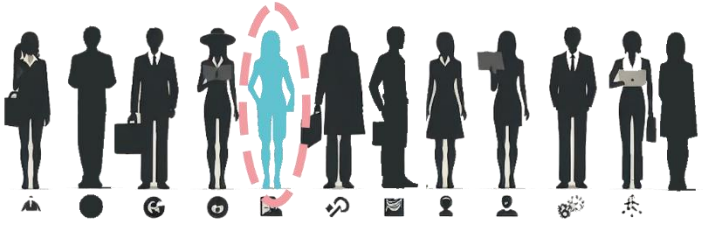
- Easy to implement („almost like not sampling at all“)
- „Quick and dirty“

Disadvantages

- Very weak external validity (weak generalizability, weak representativity)
- High risk of bias

Modal Instance Purposive Sampling
Going with the most typical respondents.

Drawing a sample based on the most typical cases (=modal cases) in the population



Defining what the modal case of the population looks like (according to e.g. age, education, income, etc.)

Advantages

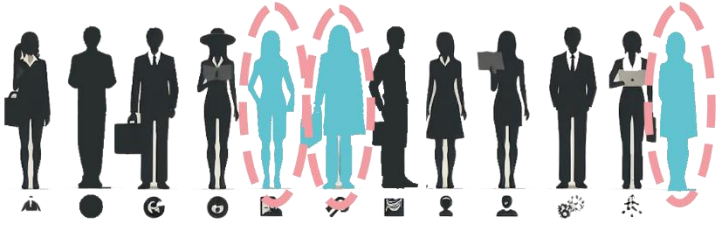
- Easily understood by audience
- When indending to measure only typical respondents

Disadvantages

- Difficulty in justifying the „typical" case (e.g. what if sampled according to age, but religion is more significant)
- Risk of bias, exclusion of diversity and minority sub-groups, results limited to modal cases

Expert Sampling
Relying on experts‘ suggestions

Drawing a sample based on experts‘ recommendations



Expert: „For answering xy research question, middle-aged women can provide the richest answers“

Advantages

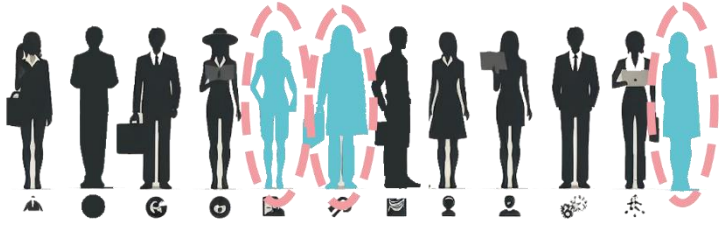
- Experts providing additional backup that supports sampling method (use in combination with other methods)

Disadvantages

- Experts can also be wrong or biased
- Hard to justify, need to legitimise expert status
- Limited external validity

Quota Sampling
Sub-group insights

Drawing a sample where you stop sampling when achieving a specific number of units for each subgroup of a population



Focus lies on sampling ideas/opinions instead of people or units – goal is not ultimate generalizability of average value

Advantages

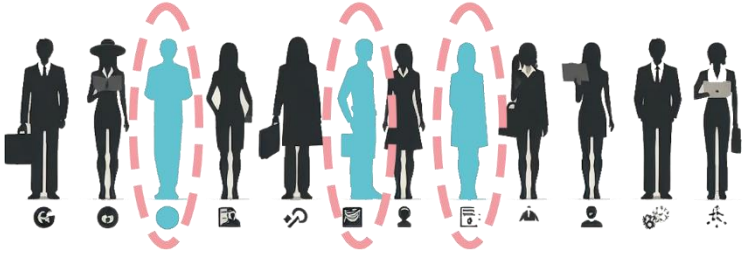
- Used for (higher) representation of sub-groups
- Allows for oversampling of minorities

Disadvantages

- Risk of distortion (depends of which units come along when)
- Complex generalizing back to population

Heterogeneity Sampling
Dealing with diversity.

Drawing a sample based on reaching the most diverse answers – prioritising heterogeneity of answers over representativity



Advantages

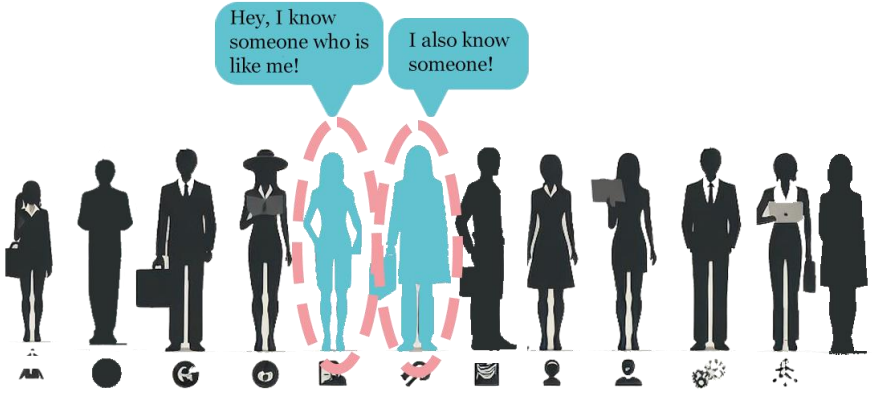
- When intending to get a broad spectrum of answers
- Allowing for oversampling of subgroups/minorities

Disadvantages

- Weak representativity, hardly/no generalizability

Snowball Sampling
Reaching out to hard-to-reach populations.

Starting with identifying units that meet the criteria for inclusion in the study and then asking them to recommend others who they know also meet the criteria



Advantages

- Method for hardly accessible populations or especially sensible topics with no records (e.g. homeless, addicts, etc.)
- Used when sampling frame is missing

Disadvantages

- Could be highly biased, distorted and one-sided sample (e.g. targetting certain „bubbles“)
- Low external validity

Sampling Methods
Overview

Probability Sampling	
Simple Random Sampling	*each potential sample has an identical chance of being chosen*
Stratified Random Sampling	*A random sample out of several homogenous sub-groups*
Systematic Random Sampling	*a sample based on a rule to select each x. element*
Cluster Random Sampling	*randomly selecting clusters and then using **each** element for the sample*
Multistage Random Sampling	*Combining several sampling techniques for different stages of sampling*

Non-Probability Sampling	
Convenience Sampling	*Randomly approached units are available or willing to participate in the study*
Modal Instance Purposive Sampling	*sample based on the most typical (=modal) cases*
Expert Sampling	*a sample based on experts' recommendations*
Quota Sampling	*stop sampling when achieving a specific number of units for each subgroup*
Heterogeneity Sampling	*a sample based on reaching the most diverse answers*
Snowball Sampling	*asking appropriate units to recommend others*

EXERCISE
Which sampling method would you use?

Match the sampling method with the scenario.

	Scenario	Resources
1	**a study on consumer preferences for a new product**	A list of 100 potential participants
2	**job satisfaction among employees in a multinational corporation**	The company has three departments: Marketing, Finance, and Operations
3	**smartphone usage patterns among university students**	Surveying students who are waiting in the campus cafeteria during lunchtime
4	**cross-cultural communication**	Interviewing expatriates who have lived in different countries

Stratified Sampling

Convenience Sampling

Expert Sampling

Simple Random Sampling

Key Takeaways

Sampling:

- **Sampling** is the process of **selecting units from a larger population**.
- Key objective of sampling is **external validity** by generalizing results back to the larger population.
- The sampling distribution will always be **a bell curve** (as long as the sample size is large enough).
- There is **always uncertainty** in an estimate, but we can quantify this uncertainty with confidence intervals (e.g., we are 95% confident that the population parameter is between X and Y).

Sampling Methods:

- There is a great variety of **different probability and non-probability sampling methods** available for researchers.
- Probability sampling methods are more **accurate and rigorous**, but non-probability sampling methods are more **feasible**.