BIOSTATISTICS

Unit 1

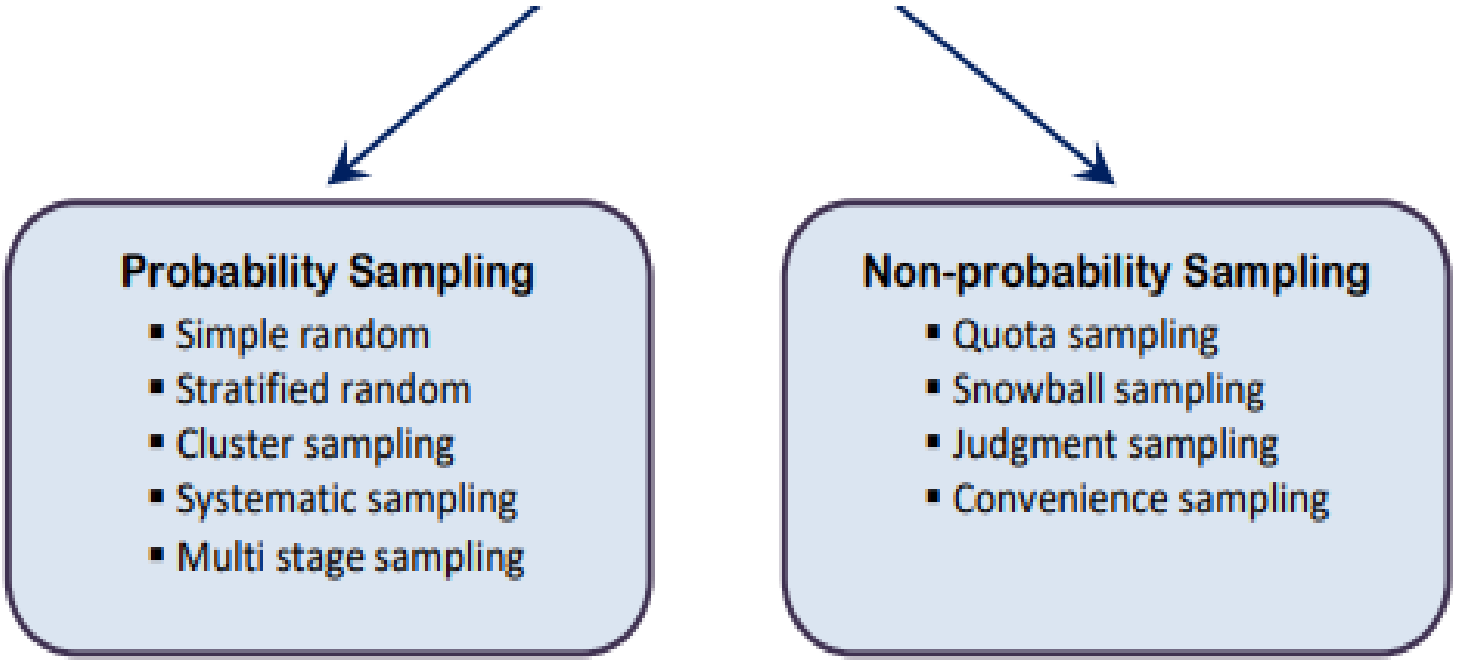
Hana Taha, PhD

BIOSTATISTICS

SAMPLING TYPES

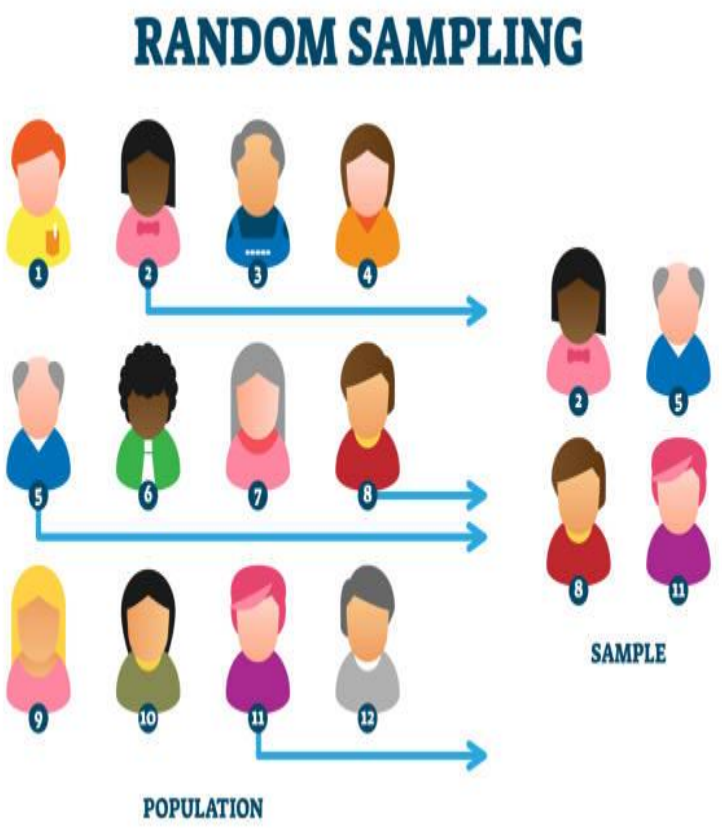
Lecture 2

Type of sampling methods



Probability Sampling Methods

- Involves random selection procedures to ensure that each unit of the sample is chosen on the basis of chance
- All units of the study population should have an equal or at least a known chance of being included in the sample
- Requires a sampling frame that lists all the study units

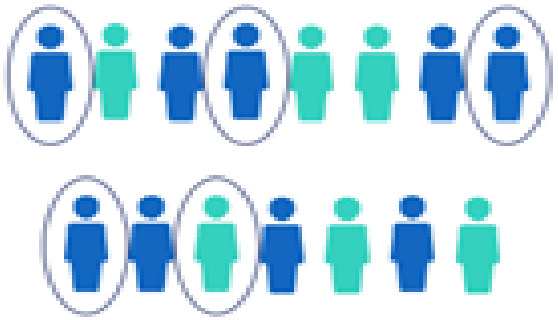


Probability Sampling Methods.

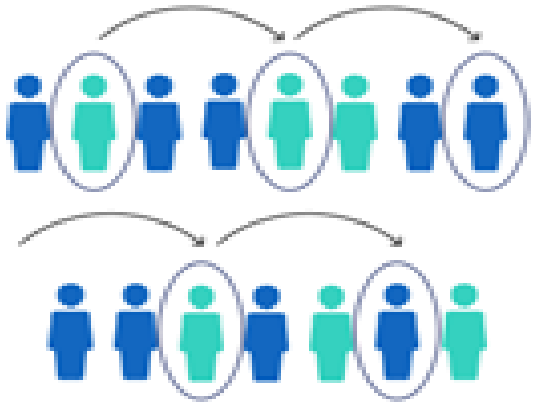
- Involves the use of random selection process to select a sample from members or elements of a populations.

  - Simple Random Sampling
  - Systematic sampling.
  - Stratified sampling.
  - Cluster sampling.
  - Multistage sampling.

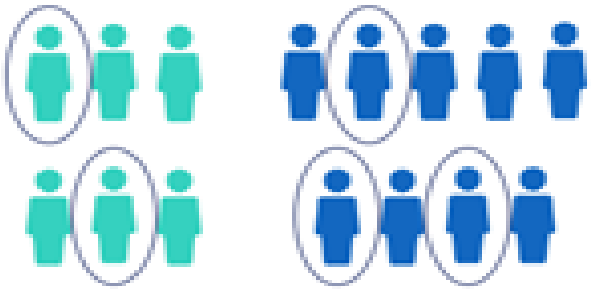
Simple random sample



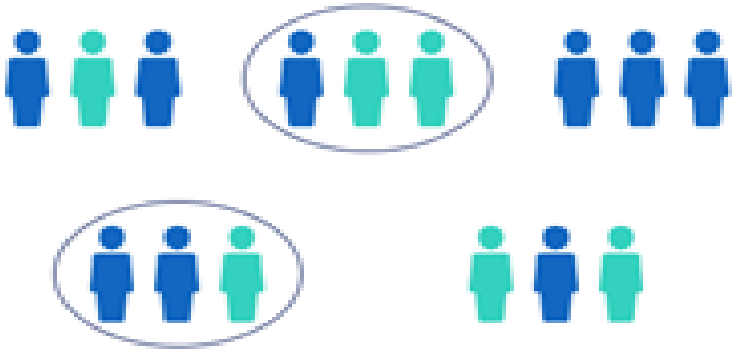
Systematic sample



Stratified sample



Cluster sample



Simple Random Sampling

- This is the simplest of probability sampling
- Make a numbered list of all units in the population
- Decide on the sample size.
- Select the required number of sampling units using the lottery method or a random number table

27798	12511	31487	23968	92052	69972
20741	65030	79887	60896	34880	80647
72357	22780	68845	07401	55229	40009
04912	18963	98752	62310	56615	37512
65671	11784	94998	49452	67552	87610
26900	84444	07973	17655	41558	29754
85865	45148	13069	40746	57146	28399
64831	27098	10675	17555	17833	72997
11063	88498	26184	71909	52135	78002

Using a random number table to make a fair decision

Pros and Cons:

- **Strong external validity :** Allows researchers to generalize results from the sample to the entire population being studied.
- **Relative speed and efficiency compared to the census:** Much faster and more efficient than collecting data from every member of the population.
- **Expensive:** Contacting a large, randomly selected group of people requires lots of resources.
- **Time consuming:** Although this method is faster than conducting a census, gathering data from a large, random sample is often slow when compared to other methods.
- **Not always possible:** Researchers may wish to study a group for which there is no organized list (sampling frame) to randomly sample from.

Example of Simple Random Sampling:

- Researchers who want to assess what Jordanians know about a particular health topic might use simple random sampling. The researchers could begin with a list of telephone numbers from a database of all cell phones and landlines in Jordan. Then randomly select numbers to call, ensuring a simple random sample.

Systematic Sampling

- Individuals are chosen from the sampling frame at regular intervals
- Ideally, we randomly select a number to tell us the starting point
  - every 5th household
  - every 10th women attending ANC
- Sampling fraction = $\frac{\text{Sample size}}{\text{Study population}}$

- Interval size= $\frac{\text{study population}}{\text{Sample size}}$

Systematic Sampling

In systematic sampling, the first individual is chosen at random, then each is chosen by a fixed interval.

Population

Sample

Every third person.

Pros and Cons:

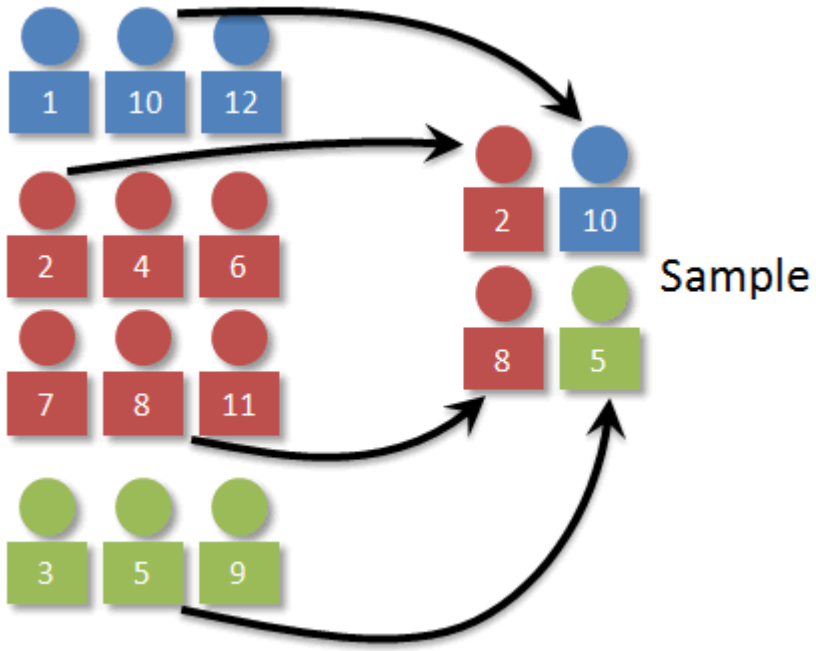
- **External validity:** Systemic sampling allows generalization from the sample to the population being studied.
- **Relative speed:** Faster than contacting all members of the population or simple random sampling.
- **Limited feasibility:** This sampling method is not possible without a list of all members of the population.

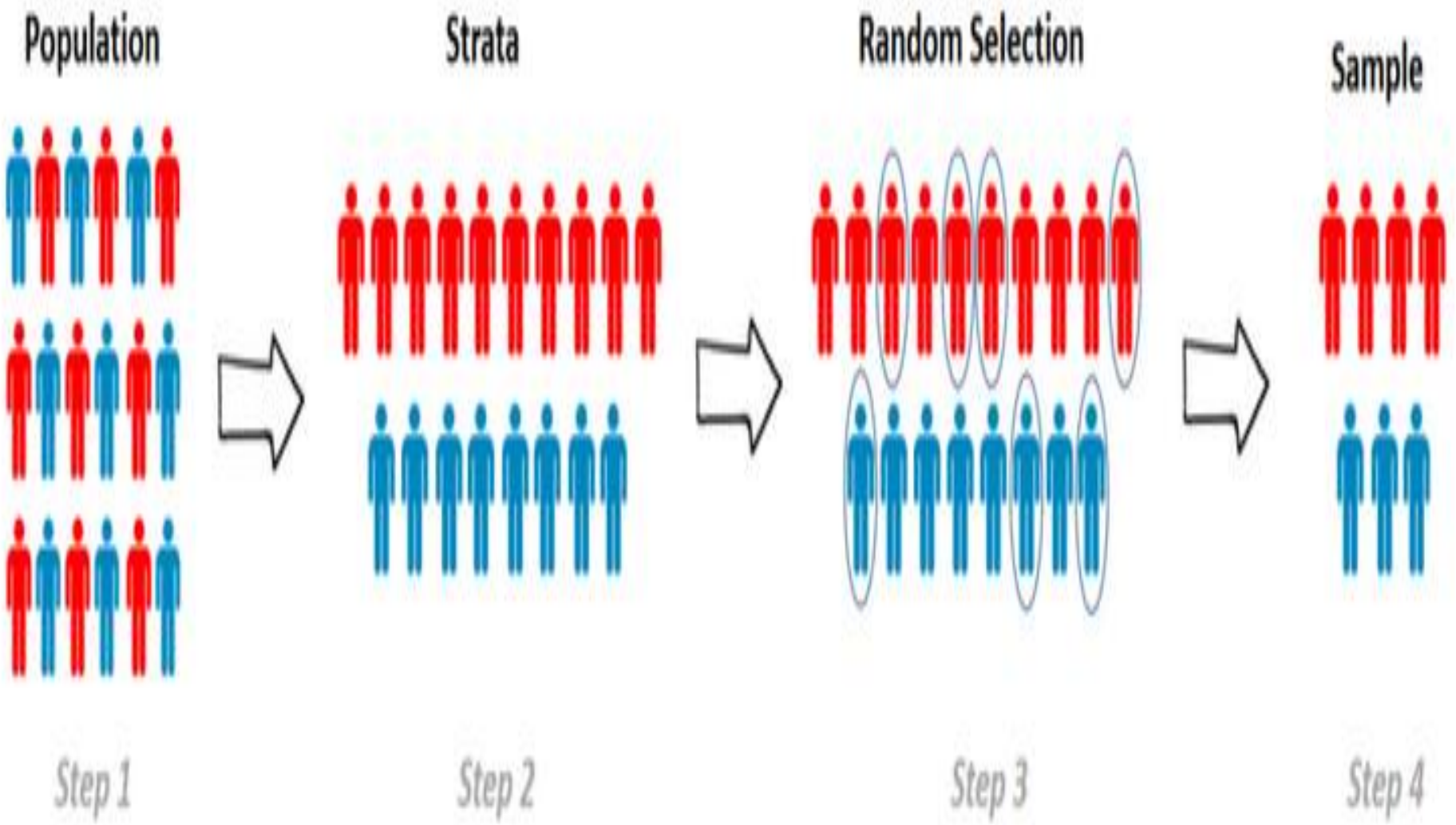
Example of Systematic Sampling

- Colleges and universities sometimes conduct campus-wide surveys to assess students' perceptions toward campus climate. The university could use systematic sampling by starting with a list of all registered students, then randomly select a starting point and an interval to sample with. Contacting every student who falls along the interval would ensure a random sample of students

Stratified Sampling

- If we have study units with different characteristics which we want to include in the study then the sampling frame needs to be divided into strata according to these characteristics
- Ensures that proportions of individuals with certain characteristics in the sample will be the same as those in the whole study population
- Random or systematic samples of predetermined sample size will have to be obtained from each stratum




Population
Strata
Random Selection
Sample
Step 1
Step 2
Step 3
Step 4

Pros and Cons:

- **External validity:** Maintains the researcher’s ability to generalize from the sample to the entire population being studied.
- **Representation:** By selecting important groups to sample from, the researchers can ensure adequate representation of small and minority groups.

An Example of Stratified Sampling:

- Researchers may be interested in assessing the attitude of university students about smoking. To ensure that members of all various colleges are adequately represented in their surveys, these researchers might use stratified sampling. In doing so, researchers would select students from these strata that is important to represent in the study for example (health colleges, Engineering, Colleges of Humanities and Social Sciences) and then randomly sample students who belong to each of these strata.
- By using this technique, the researchers can ensure that all the students are adequately represented in the sample while maintaining the ability to generalize their results to the larger population.

Cluster Sampling

- Selection of study units (clusters) instead of the selection of individuals
- All subjects/units in the cluster who meet the criteria will be sampled.
  - Clusters often geographic units
  - e.g. schools, villages etc
- Usually used in interventional studies
  - E.g. assessing immunization coverage
- Advantages
  - sampling frame is not required in this case
  - Can be costly because the study population can be scattered over a large area

Pros and Cons:

- External validity: The random nature of selecting clusters allows researchers to generalize from the sample to the entire population being studied.
- Speed: Faster and more efficient than sampling all people in the population.
- Cost reduction because the clusters division is natural
- Not always possible if there is no organized list from which to randomly select participants from some clusters.

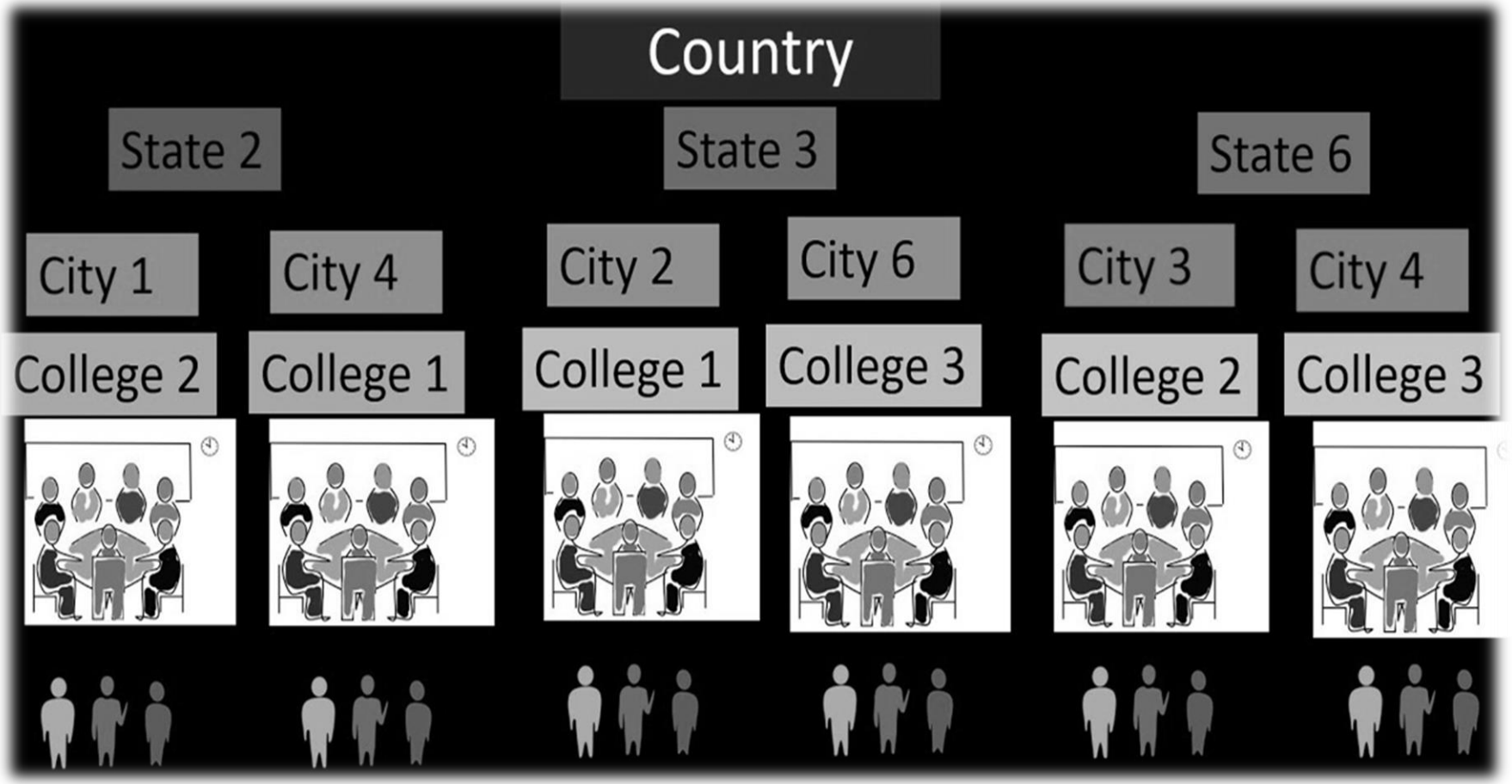
Multi-stage sampling (also known as multi-stage cluster sampling)

- Multistage sampling is a version of cluster sampling.
- Multistage sampling begins when researchers randomly select a set of clusters or groups from a larger population.
- It is more complex form of cluster sampling because it contains two or more stages in sample selection.
- In multi-stage sampling clusters are selected in several stages to make primary data collection more manageable.

Multistage Sampling

- Multi-stage sampling is not as effective as true random sampling; however, it addresses certain disadvantages associated with true random sampling such as being overly expensive and time-consuming.
- Does not require an initial sampling frame of whole population
- Require sampling frames of final clusters
- Applicable to community-based studies e.g. interviewing people from different villages selected from different districts, selected from different provinces.

Multi-stage sampling

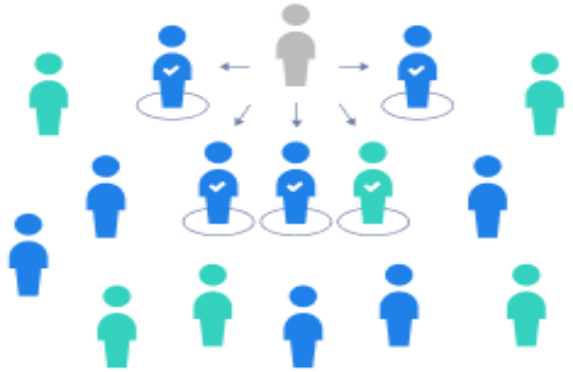


Nonprobability Sampling Methods

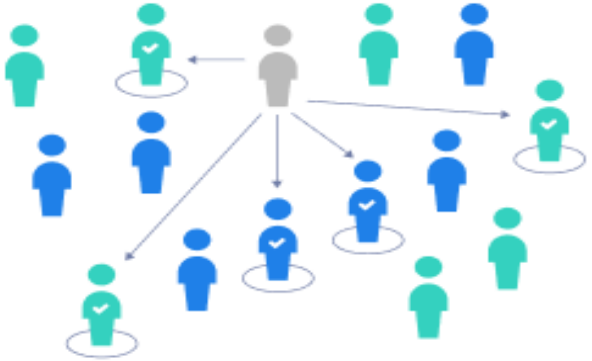
Nonprobability sampling: the sample elements are chosen from the population by nonrandom methods.

- More likely to produce a biased sample than the random sampling.
- This restricts the generalization of the study findings.
- Most frequent reasons for use of nonprobability samples involve convenience and the desire to use available subjects.

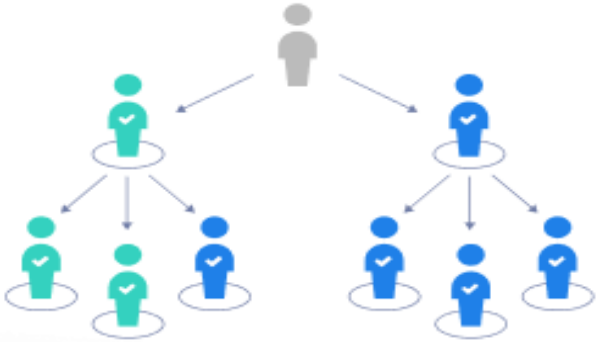
Convenience sample



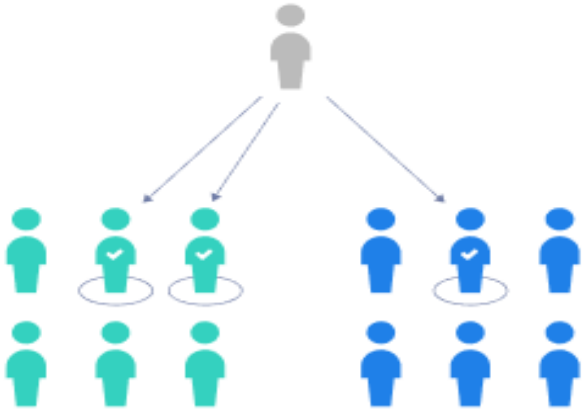
Purposive sample



Snowball sample



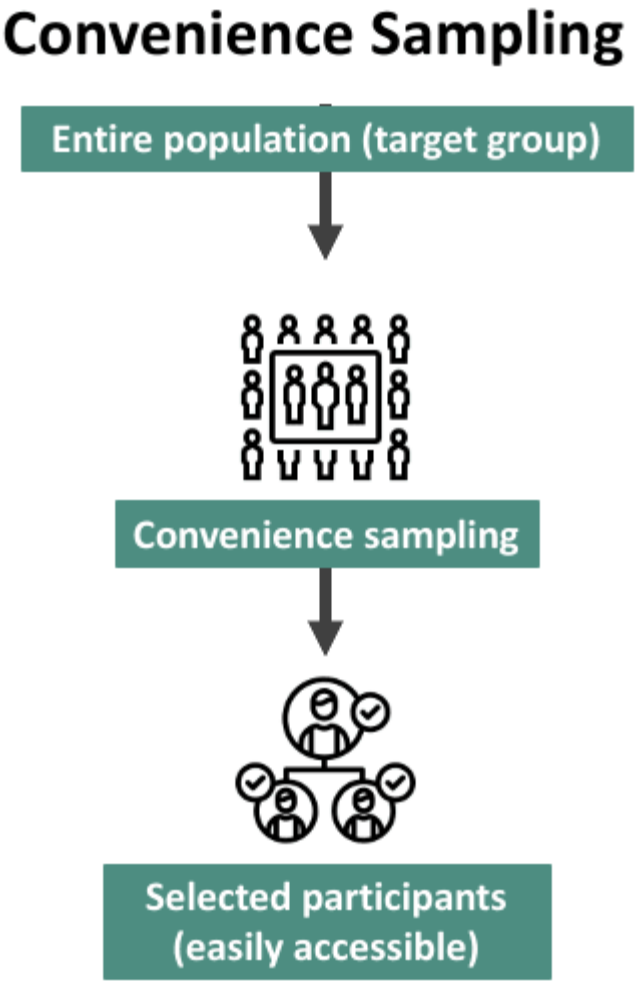
Quota sample



Convenience sampling (Accidental or incidental sampling):

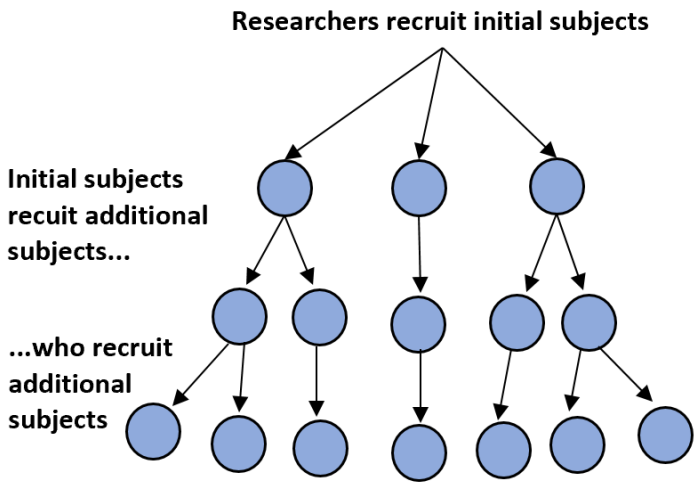
- Convenience samples are often based on who is easy for the researchers to recruit for the study.
- People may or may not be representative of the population.
- Most frequently used in health research

- Advantages:
  - Saves time and money



Snowball sampling

Snowball sampling begins when researchers contact a few people who meet a study's criteria. After those people complete the study, the researchers ask each person to recommend a few others who also meet the study criteria. By building on each participant's social network, the data collection will snowball until the researchers reach enough people for their study.

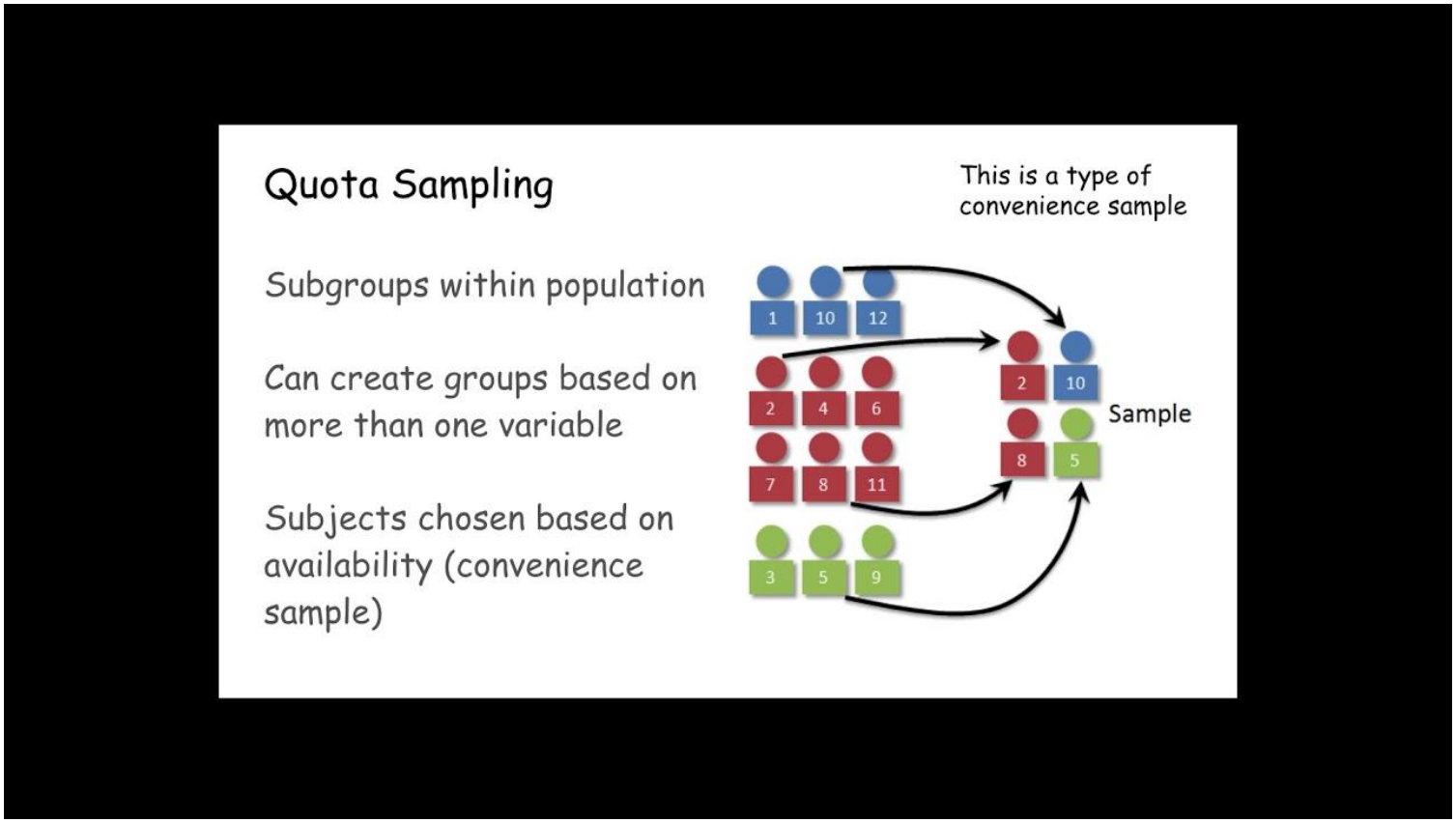


Pros and Cons:

- By drawing on people's social networks, snowball sampling can be an effective way to study hard-to-reach groups. Once researchers gain the trust of a few members of the group, those people can help the researchers recruit other people.
- Ability to reach small or stigmatized groups in topics of research where the subjects are reluctant to make their identity known, Drug users, Aids patients, etc.
- Non-random: A snowball sample will likely provide results that are hard to generalize beyond the sample studied.
- Slow: Because snowball sampling relies on each participant to recommend others, the data collection process is typically slow when compared to other methods.

Quota sampling

- When researchers decide to use quota sampling, they identify subsets of the population that are important to represent and then sample participants within each subset.



Advantages

- Easy and quick results
- Saves time, money and energy

Disadvantages

- Not representative of population
- Cannot produce reliable results

Purposive sampling (handpicking, judgmental)

- Subjects are chosen because they are typical or representative of the accessible population, or because they are experts (more knowledgeable) in the field of research topic.

- Qualitative researchers use Purposive sampling