



YOSH SOTSIOLOGLAR uchun yozgi maktab

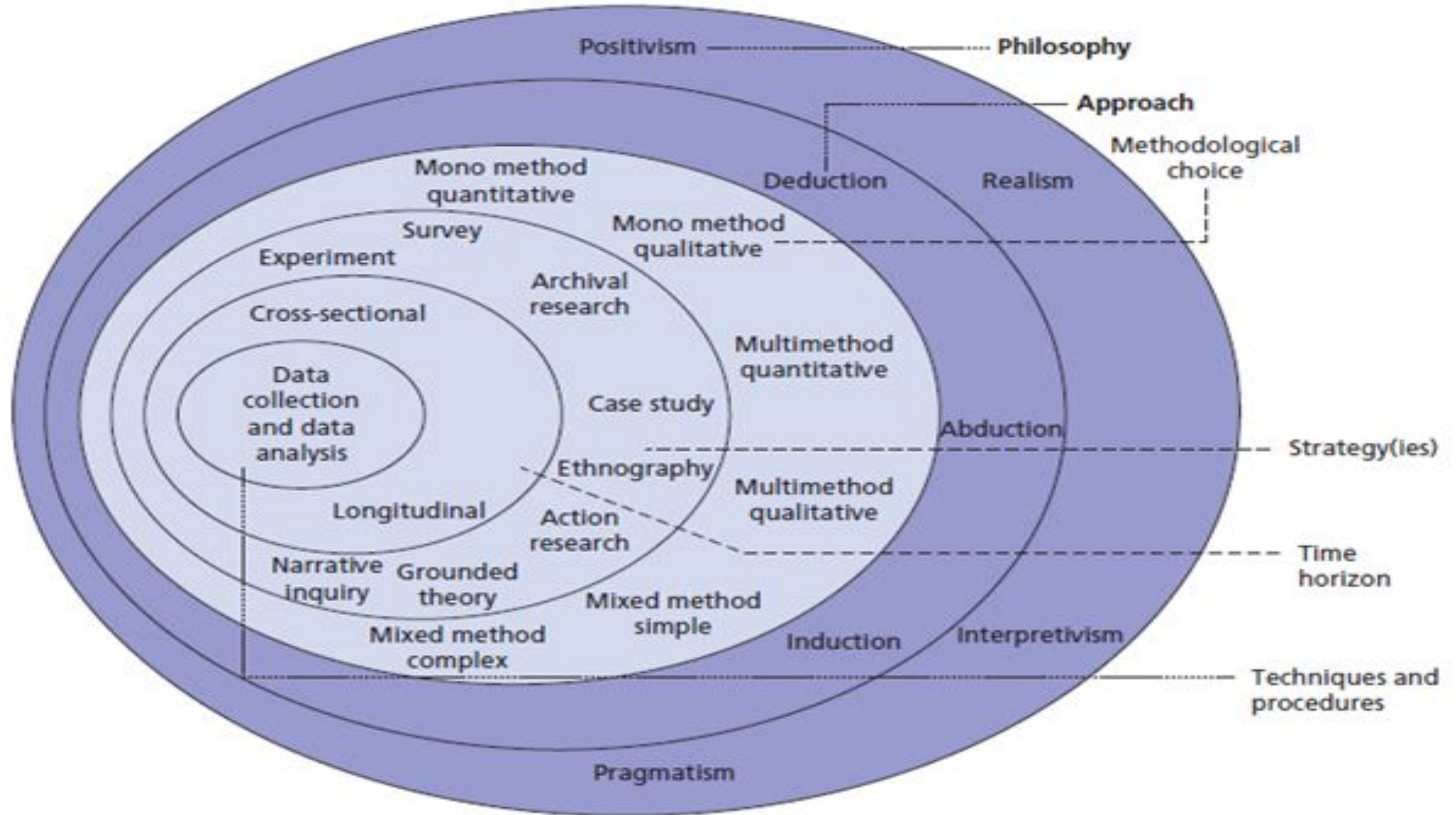


Sample & Sampling Techniques

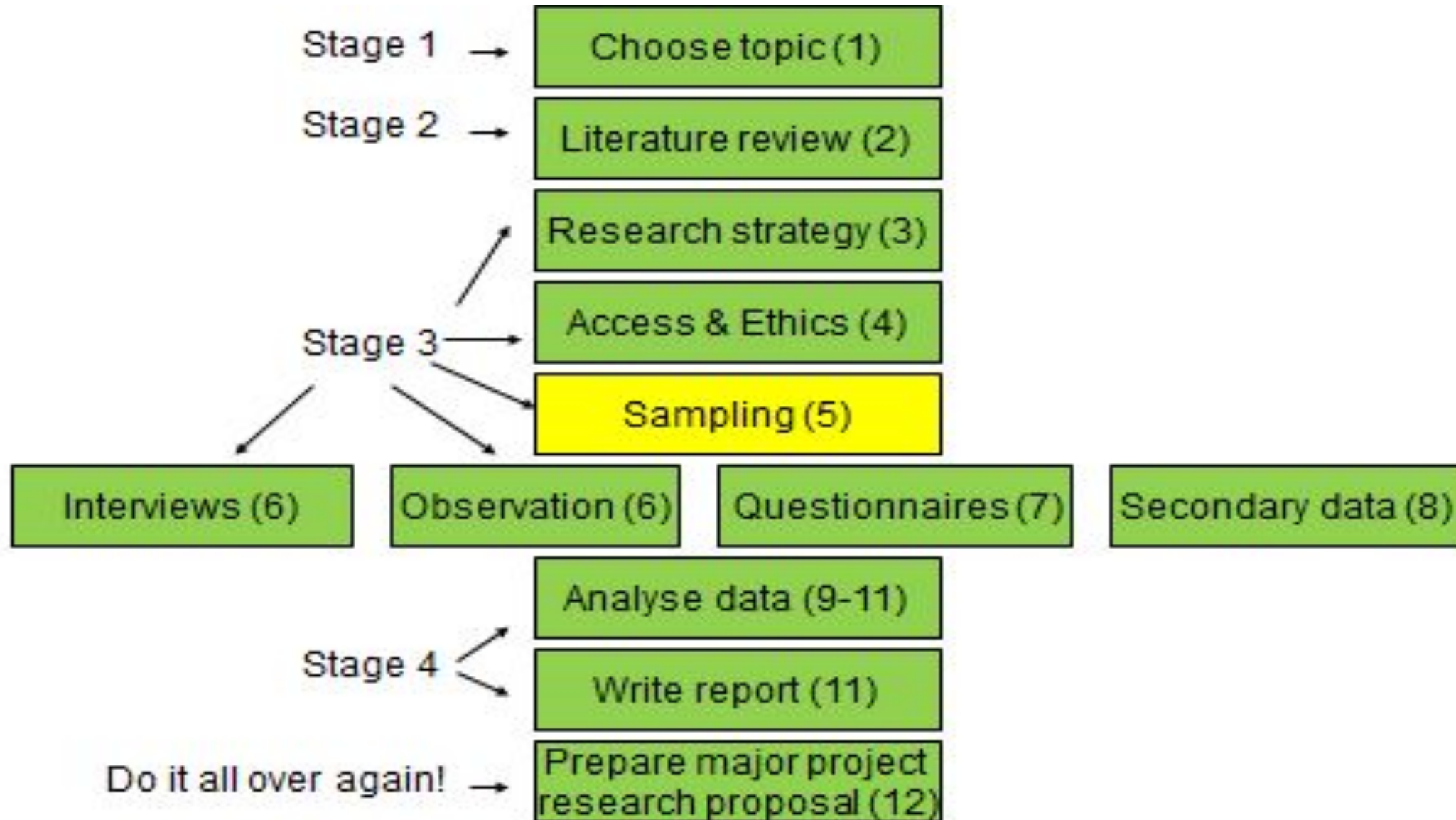
Day 3: August 23, 2023

Dr. Nargiza Alimukhamedova

“Research Onion” and role of sampling:



Sampling in stages of research:

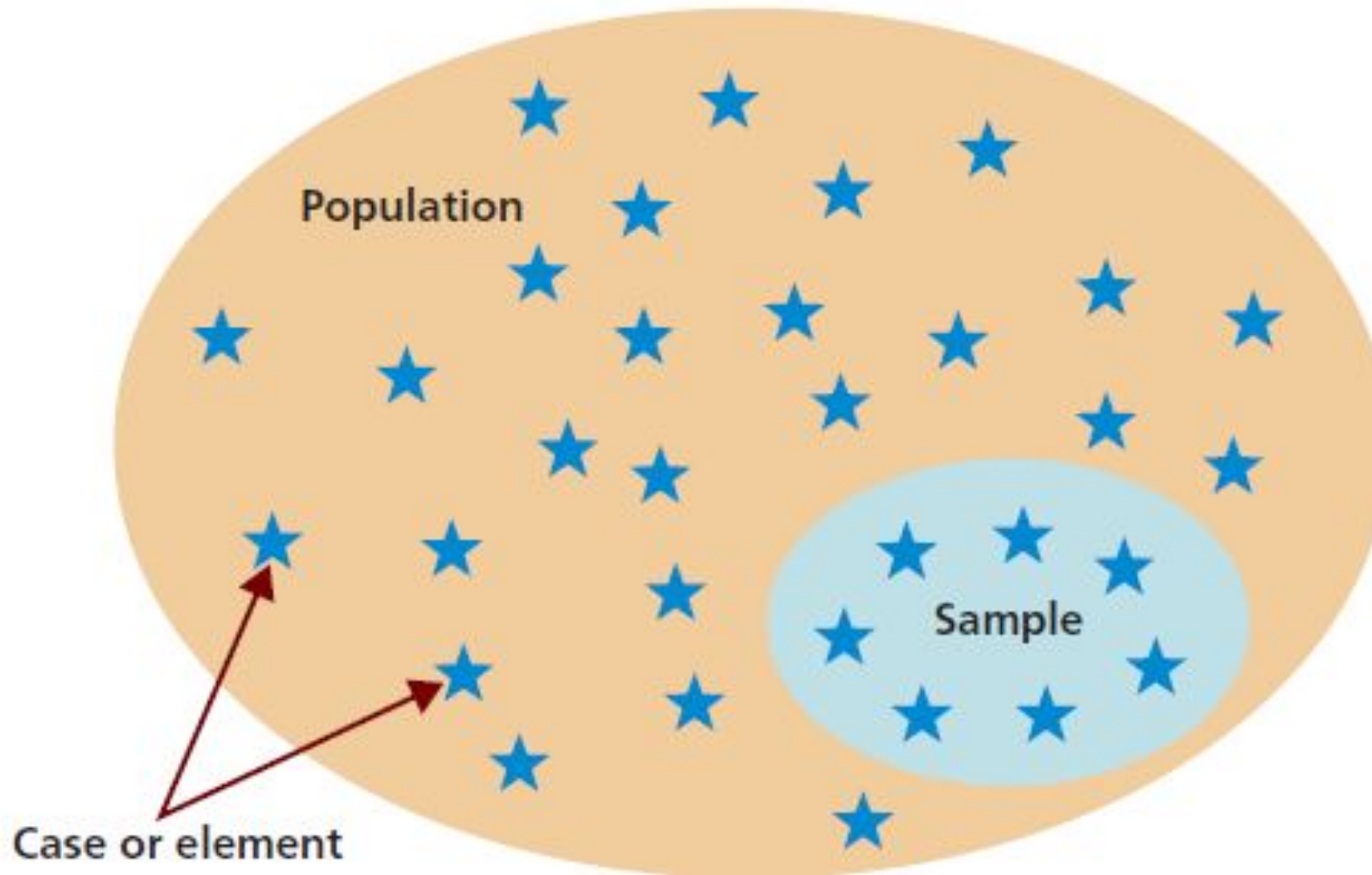


What is a Sample ?

Sample is a subset of the population. By studying the sample, the researcher should be able to draw conclusions that are generalizable to the population of interest.

Sampling is an act, process, or technique of selecting a suitable sample, or a representative part of a population for the purpose of determining parameters or characteristics of the whole population.

What is a Sample ?



Population & Sample

<https://www.youtube.com/watch?v=eIZD1BFfw8E>

Survey for job prospects



That's not the population!
The population of interest includes not only
the students on campus but also the ones at

365° DataScience

Population vs Sample

Importance of Sampling:

1. We can't always survey EVERYONE in the study

Example: survey all 3.500 students in WIUT; survey 10.000 employees of GM

2. The magic in research methods: you can do sampling instead. If you do proper sampling, then all finding based on your sample would be also generalizable to whole population.




3. Sampling saves you time and work

4. Sampling enables your research to have relevance and meaning

How to calculate Sample Size ?

- How many people do you need to take your survey?
- Even if you're a statistician or math expert, determining survey sample size is not an easy task.

Calculate your sample size

Population size 	Confidence level (%) 	Margin of error (%) 
<input type="text" value="1000"/>	<input type="text" value="95"/> 	<input type="text" value="5"/>

Sample size

278

Sampling techniques:

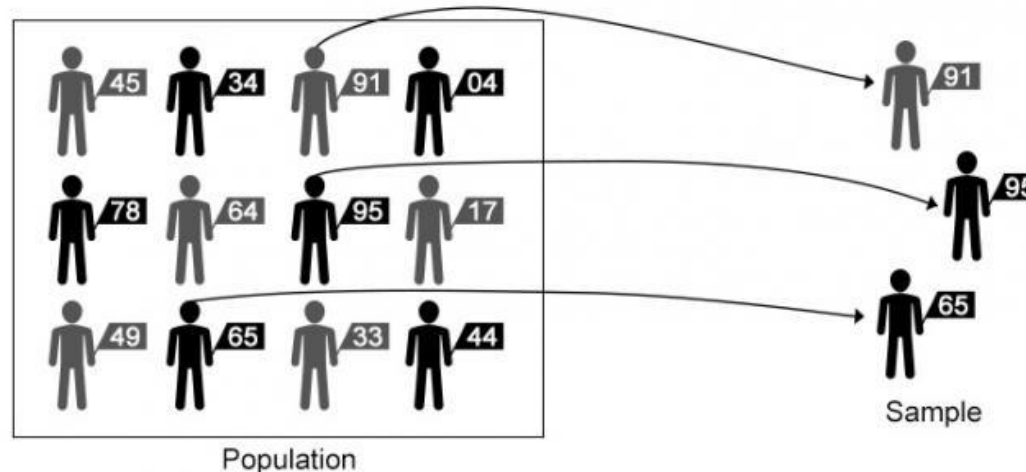
There are two types of selecting samples:

- I. Probability sampling**
- II. Non-probability sampling**

I. Probability sampling: Overview

Probability sampling is the way of sampling in which each item in the population has an equal chance (this chance is greater than zero) for getting selected is called probability sampling.

- Chance of each case/element being selected is known and usually equal.
- You can estimate statistically the characteristics of the whole population.
- Probability Sampling uses lesser reliance over the human judgment which makes the overall process free from over biasness.



Probability Sampling

I. Probability sampling: Overview

Advantages:

- 1) **Cost Effective**
- 2) **Involves lesser degree of judgment**
- 3) **Comparatively easier way of sampling**
- 4) **Less time consuming**
- 5) **Can be done even by non-technical persons**
- 6) **Sample representative of population**

Disadvantages:

- 1) **Chances of selecting specific class of samples only.**
- 2) **Redundant and monotonous work**

I. Probability sampling: steps

Step 1: Choose a sampling **frame**

Step 2: Decide on sample **size**

Step 3: Select an appropriate sampling **technique**

Step 4: Check that the sample is **representative of the population**

Step 1: Choose a sampling frame

Sampling frame is a list of all the cases in the population from which sample is drawn.

Sampling frame must be:

- 1) Relevant/current
- 2) Complete
- 3) Free from bias
- 4) Precise and accurate

Example: list of students from; list of employees from HR department

Minimum sample size:

Minimum sample size for different sizes of population:

Margin of error				
Population	5%	3%	2%	1%
50	44	48	49	50
100	79	91	96	99
200	132	168	185	196
400	196	291	334	384
500	217	340	414	475
1,000	278	516	706	906
2,000	322	696	1091	1655
5,000	357	879	1622	3288
10,000	370	964	1936	4899
100,000	383	1056	2345	8762
1,000,000	384	1066	2395	9513

Saunders et al. (2003)

I. Probability sampling (representative)

Step 1: Choose a sampling frame

Step 2: Decide on sample size

Step 3: Select an appropriate sampling technique

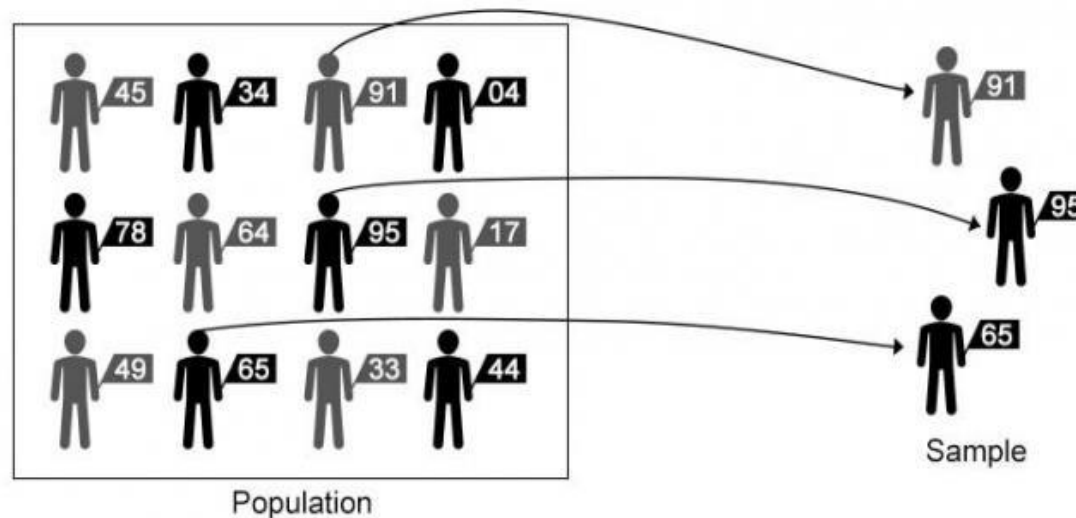
There are 5 probability sampling techniques:

- (a) Simple random
- (b) Systematic random
- (c) Stratified random
- (d) Cluster

(a) Simple random sampling

Definition: It is a process of assigning the random numbers to the elements of the population and selecting some of them by way of certain specific rule (like highest among the local group/row lowest among the group/row etc).

How to do sampling? Number the cases and then select at random using tables or computer. There are ready programs of “Random Number Generators”



Simple Random Sampling

How to generate random numbers?

You can generate random numbers either (1) using on-line ready service or using simple (2) Excel:



Home Games Numbers Lists & More Drawings Web Tools Statistics Testimonials Learn More Login

RANDOM.ORG

Search RANDOM.ORG Search

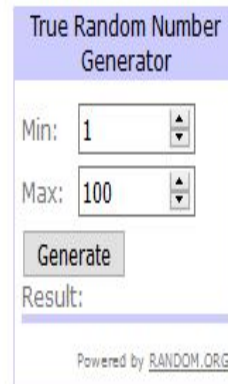
True Random Number Service

Do you own an iOS or Android device? [Check out our app!](#)

What's this fuss about *true* randomness?

Perhaps you have wondered how predictable machines like computers can generate randomness. In reality, most random numbers used in computer programs are *pseudo-random*, which means they are generated in a predictable fashion using a mathematical formula. This is fine for many purposes, but it may not be random in the way you expect if you're used to dice rolls and lottery drawings.

RANDOM.ORG offers *true* random numbers to anyone on the Internet. The randomness comes from atmospheric noise, which for many purposes is better than the pseudo-random number algorithms typically used in computer programs. People use RANDOM.ORG for holding drawings, lotteries and sweepstakes, to drive online games, for scientific applications and for art and music. The service has existed since 1998 and was built by [Dr Mads Haahr](#) of the [School of Computer Science and Statistics](#) at [Trinity College, Dublin](#) in Ireland. Today, RANDOM.ORG is operated by [Randomness and Integrity](#)



True Random Number Generator

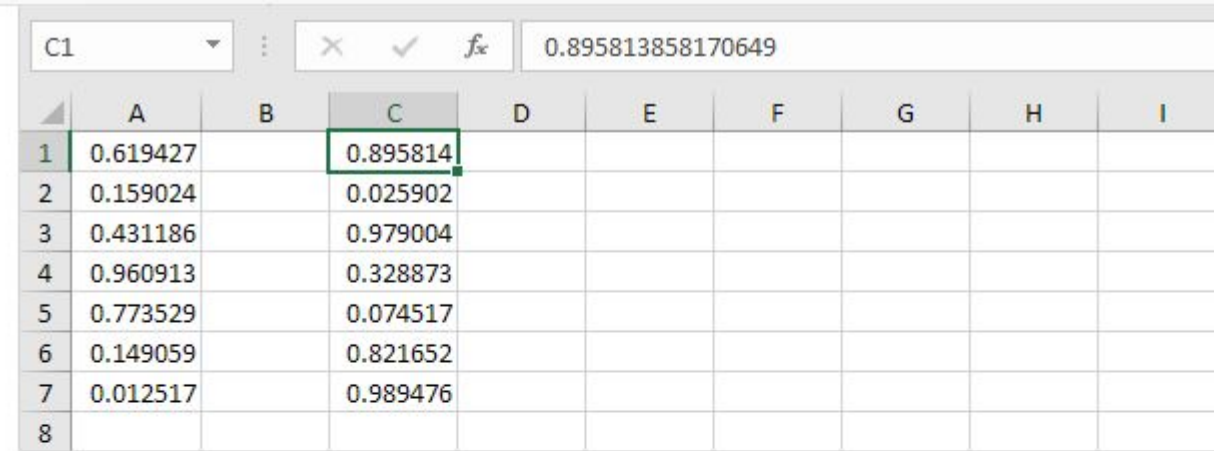
Min:

Max:

Result:

Powered by [RANDOM.ORG](#)

<https://www.random.org/>

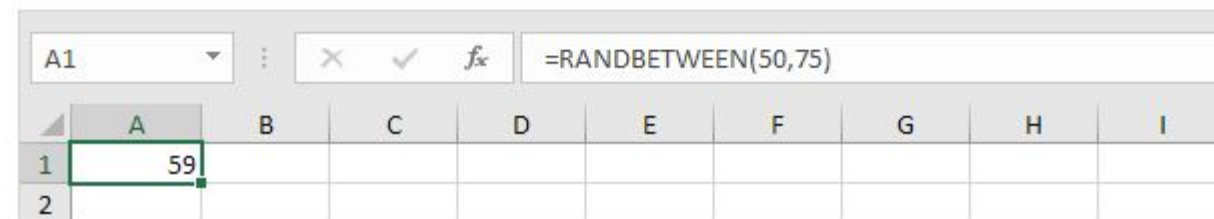


	A	B	C	D	E	F	G	H	I
1	0.619427		0.895814						
2	0.159024		0.025902						
3	0.431186		0.979004						
4	0.960913		0.328873						
5	0.773529		0.074517						
6	0.149059		0.821652						
7	0.012517		0.989476						
8									

Randbetween

The **RANDBETWEEN** function generates a random whole number between two boundaries.

1. Select cell A1.
2. Type **RANDBETWEEN(50,75)** and press Enter.



	A	B	C	D	E	F	G	H	I
1	59								
2									

<https://www.excel-easy.com/examples/random-numbers.html>

(a) Simple random sampling

Advantages:

- 1) Better chances that the sample represents the whole population.
- 2) Can be concluded in shorter time duration.
- 3) Costs less money.
- 4) Involves lesser degree of judgment.
- 5) Comparatively easier way of sampling.
- 6) Can be done even by non-technical persons too.

Disadvantages:

- 1) Risks of selecting samples from a few variations only.
- 2) Redundant and monotony.

(a) Simple random sampling

When simple random sampling is applicable?

- **With samples of a few hundred**
- **When frame is easily accessible (computer)**
- **When travel is not an issue (if you need to speak face to face with participants)**

Example: You conduct study on Blended Learning among WIUT students:

Step 1. You take the list of all WIUT students from Registrar.

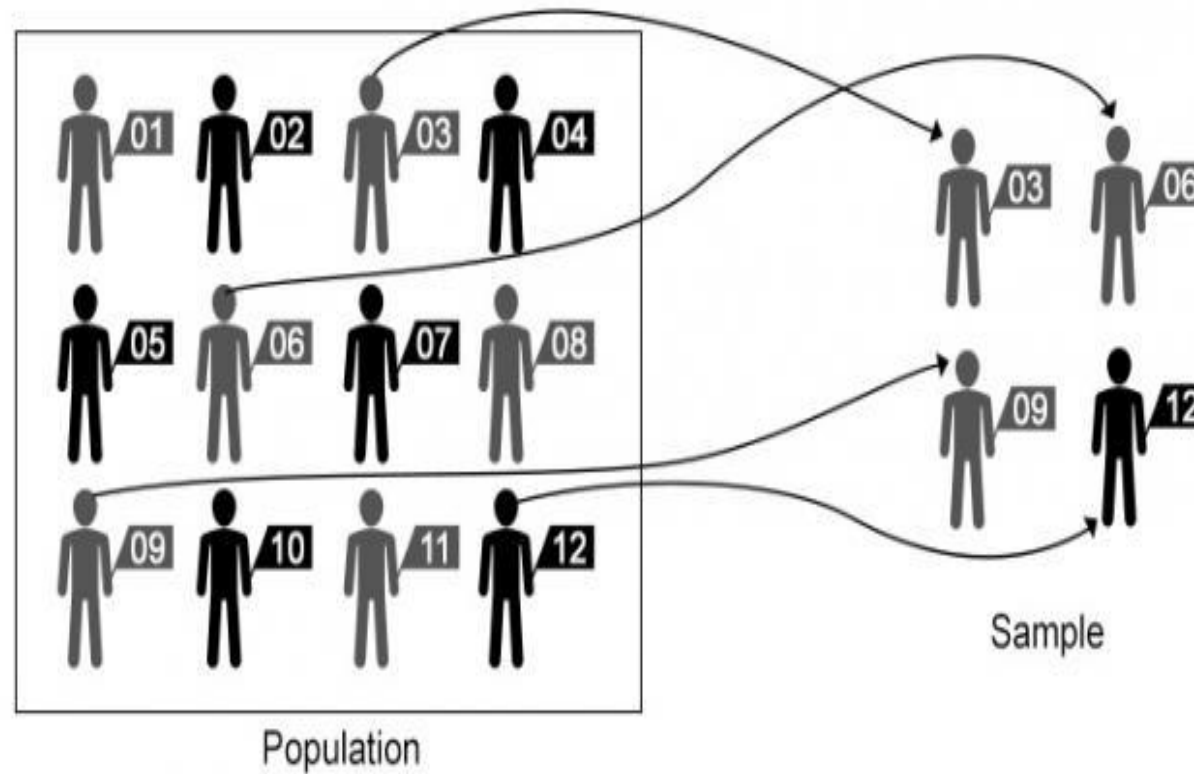
Step 2. You generate random numbers (either online or using Excel)

Step 3. You apply randomly generated numbers to select cases into your sample

Step 4. You continue selecting cases until you reach your sample size

(b) Systematic random sampling

Definition: Systematic sampling involves you selecting the sample at regular intervals from the sampling frame.



Systematic Random Sampling

(b) Systematic random sampling

How to do Systematic random sampling?

- 1) Determine the size of the population.
- 2) Determine the sample size (number of samples to be taken).
- 3) Calculate the number of items in the population to be represented by each sample.
- 4) Number of items in the population to be represented by each sample
 $(n) = \text{Population Size} \div \text{Sample Size}$.
- 5) Select any item as the first sample and choose next samples which comes at nth position after the previous one. This is also know as a **WALK SIZE** which is for example every 5th or every 7th
- 6) Collect the samples and draw an appropriate conclusion.

(b) Systematic random sampling

Advantages:

- 1) Easier than Stratified Random Sampling and more effective than Simple Random Sampling.
- 2) Cost Effective.
- 3) Avoids Judgments.
- 4) Less Time Consuming.
- 5) Higher degree of Control.

Disadvantages:

- 1) Size of the population may not be known before the sampling starts.
- 2) Assumption that the population is uniform may not always prove to be true.

(b) Systematic random sampling

When it is applicable?

- When the list does not have periodic patterns (i.e. age, gender, alphabetical order)
- Unlike random it works with small samples
- Like random, not good if travel is an issue
- When you need to easily explain sampling

Example: You conduct study on study mode during covid among students:

Step 1. You take the list of all students.

Step 2. You determine the WALK SIZE, i.e. every 5th or every 7th

Step 3. You choose one random number, and then using the WALK SIZE select cases into your sample

Step 4. You continue selecting cases until you reach your sample size

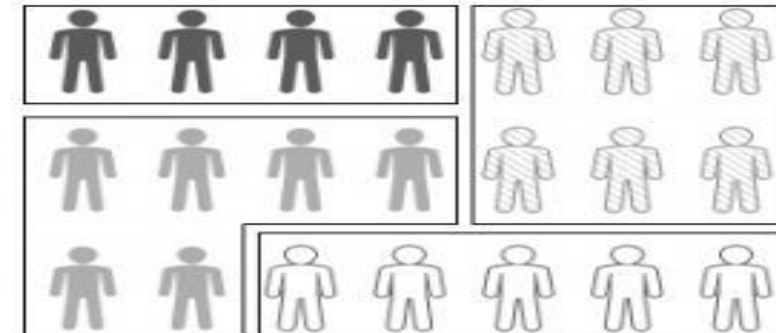
(c) Stratified random sampling:

Definition: Stratified random sampling is a modification of random sampling in which you divide the population into two or more relevant and significant strata based in a one or a number of attributes.

What is strata ? In a simple way strata means a layer. In sociology, strata implies a level or grade of a people or population with reference to social position, education, etc.: the lowest stratum of society.



Random Population



Stratified Population

Stratified Random Sampling

(c) Stratified random sampling:

Advantages:

- 1) provides better precision as it takes the samples proportional to the random population.
- 2) helps minimizing the biasness in selecting the samples.
- 3) ensures that no any section of the population are underrepresented or overrepresented.
- 4) As this method provides greater precision, greater level of accuracy can be achieved even by using small size of samples. This saves resources.

Disadvantages:

- 1) requires more administrative works as compared with Simple Random Sampling.
- 2) It is sometimes hard to classify each kind of population into clearly distinguished classes.
- 3) can be tedious and time consuming job to those who are not keen towards handling such data.

(c) Stratified random sampling:

How to do sampling?

- Divide the population into two or more relevant strata –subsets in sampling frame (age, gender, salary, department, region).
- Carry out random or systematic sampling on each strata.

When it is applicable?

- When list has clear patterns/strata (company department – will accurately reflect all departments)
- You can have more than one characteristic

Example: You conduct study on study mode during covid among WIUT students:

Step 1. You divided all students into Levels (3,4,5), Courses (BA, Econ, BIS, Law), etc. □ they become strata !

Step 2. You perform simple random or systematic random sampling on each strata

Step 3. You continue selecting cases until you reach your sample size

(d) Cluster random sampling:

Definition: this random sampling technique implies choosing randomly clusters in the first stage and then performing random sampling inside clusters

What is cluster ? It is a number of things of the same kind, growing or held together. In sociology it mainly refer to people residing in some location. For example,

Examples of cluster include people living in various regions, districts; or special group of minorities, gypsies, ethnic groups etc.



(d) Cluster random sampling:

How to do sampling?

- **Divide your frame into groups (clusters) using a characteristic as in stratified sampling**
- **Number each of the clusters and not individual cases**
- **Select clusters using a form of random sampling**
- **The sample then consists of all cases in these clusters**
- **The greater the number of clusters the greater the representativeness of the sample.**

When it is applicable?

- **Where travel is an issue**
- **When representativeness is less important**

Example: You conduct study on taxi use in Tashkent. You have a limited time and have to cover all districts of Tashkent. You decided to do cluster sampling.

Step 1. You choose randomly 4 clusters, i.e. Chilanzar, Mirobod, Yunosobod, Parkent. These clusters should not be overlapping geographically.

Step 2. You perform simple random or systematic random sampling in each cluster

I. Probability sampling (representative)

Step 1: Choose a sampling **frame**

Step 2: Decide on sample **size**

Step 3: Select an appropriate sampling **technique**

Step 4: Check that the sample is **representative of the population**

- This can be done by comparing it to other data sources.
- Compare your sample (no. of males/females, subject, level) with the university statistics.
- Can conduct simple non-parametric tests on equality of sample distributions (i.e. Kolmogorov-Smirnov test)
- Always state if you think your sample is not fully representative for whatever reason.

II. Non-probability sampling

I. Probability sampling (representative)

- Chance of each case/element being selected is known and usually equal.
- You can estimate statistically the characteristics of the whole population.

II. Non-probability sampling

- Chance of each case/element being selected is not known !
- You may be able to generalize about the whole population but not statistically !

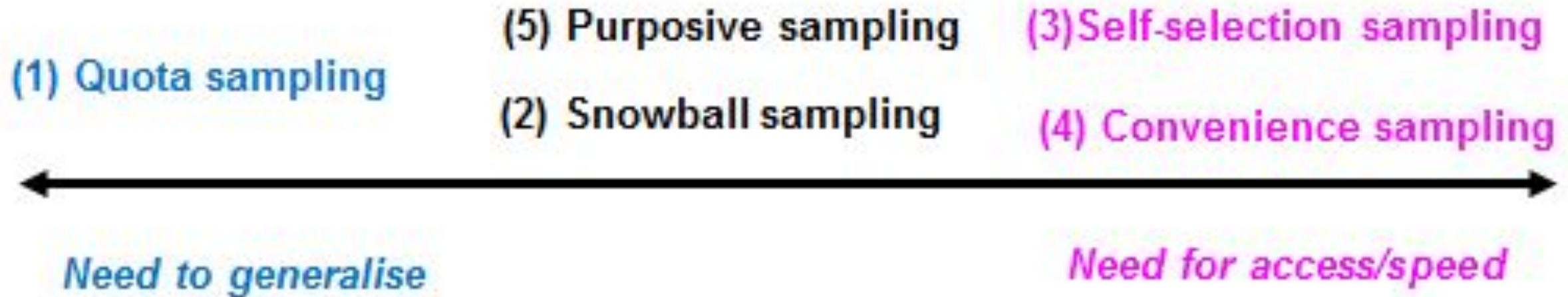
II. Non-probability sampling

Important features of non-probability sampling techniques:

- Non-probability sampling is often associated with case study research design and qualitative research.
- Usually non-probability sampling techniques are associated with higher response rates.
- Cases are not selected at random so we cannot know the probability a case will be included in the sample
- Sometimes probability sampling is not practical: limited resources or no sampling frame, or need to select a particular case because of the nature of your research question (particularly common with qualitative data)

II. Non-probability sampling

There are 5 types of non-probability sampling techniques. They could be grouped as follows based on two extremes: when issue of (a) generalization is more important or when you have to (b) worry more about speed and access to respondents



(a) Quota sampling

Definition: It is entirely non random and it is normally used for interview surveys. It is based on the premise that your sample will represent the population as the variability in your sample for various quota variables is the same as that in population. Quota sampling is therefore a type of stratified sample in which selection of cases within strata is entirely non-random.



Symbol	Age Group	No.
	11-21 Years	11
	22-31 Years	16
	32-41 Years	15
	42-51 Years	18
Total	11-51 Years	60

Quota Sampling

(a) Quota sampling

Important stages for performing quota sampling:

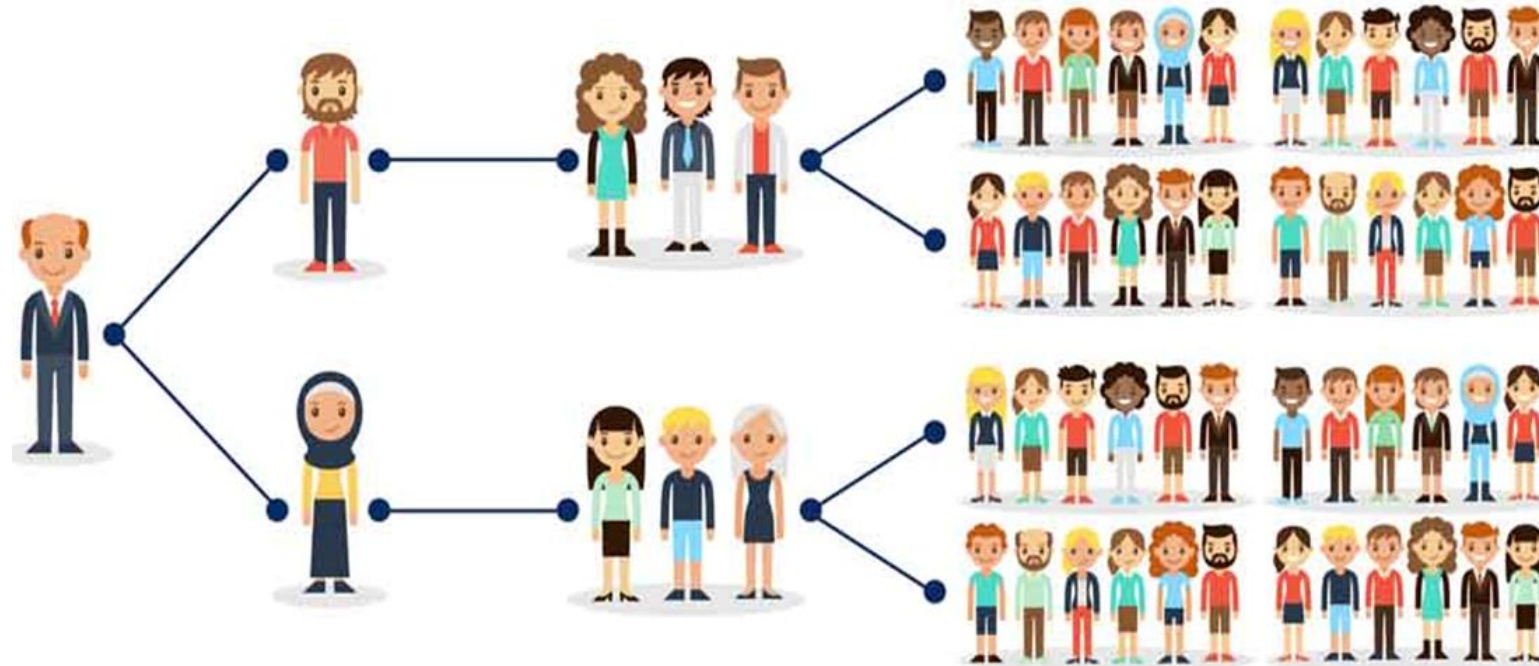
1. Divide the population into specific groups.
2. Calculate a quota for each group based on relevant and available data.
3. Give each interviewer an 'assignment', which states the number of cases in each quota from which they must collect data.
4. Combine the data collected by interviewers to provide the full sample.

When quota sampling is used ?

Usually it is used for interview surveys with large populations. You use it when there is no population frame available and where there is other data that can help you assign quotas. Helps overcome bias (e.g. pensioners)

(2) Snowball sampling:

- This sampling technique takes samples from the referrals of the previous samples. In other words, this process takes samples which are either referred by previous samples or somehow related with them.
- If you are conducting a survey regarding the popularity of a tourist destination, we can add the persons as the samples who were with the previously selected persons.



(2) Snowball sampling:

- **Snowball sampling technique makes contact with one case, ask them for other cases and keep on following up recommended cases.**
- **Snowball sampling technique is used when it is difficult to identify the sample but the cases will know each other. For example managers who have had careers in the army.**

Example: You conduct study on top management and CEOs for their career patterns

Step 1: You find access for interview for one CEO. Your friend's father is CEO 😊

Step 2: First CEO then recommend his/ her colleagues ☐ you gain access to further CEOs

(3) Self-selection sampling:

Definition: self-selection sampling technique is when cases (participants) propose themselves to take part – volunteers.

When it is used ?

- **If you want people who have strong views on a subject (so will not be representative of the population)**
- **Also useful if you are having difficulty finding cases.**

Example 1: survey work productivity of telecommuters.

Example 2: you create Google survey to study donation patterns. Therefore only those who have done donations, will be responding.

(4) Convenience sampling:

Definition: choosing those cases which are easiest and fastest to gain access to

When it is used ?

- **When there is little time and resources**
- **When there is need for speed**
- **When representativeness of sample is not a priority, or the population is homogeneous, this is a pilot study**

Example 1: interview randomly people in the shops for their preference of TV channels

Example 2: interview randomly customers in Korzinka supermarket for customer satisfaction

(5) Purposive sampling:

Definition: deliberately choosing the sample on the basis of it being particularly relevant for your research question.

When it is used ?

This sampling technique is mainly used for small samples – e.g. case study

There are 5 sub-categories of purposive sampling technique:

(5) Purposive sampling:

There are 5 sub-categories of purposive sampling technique:

- 1) **Typical case:** to give an example to the general reader. Example: used for illustrative purposes
- 2) **Extreme case:** based on outlier. Example: study only best-performing companies
- 3) **Homogenous:** choosing cases that are of the same quality. Example: study only BA students
- 4) **Heterogeneous:** choosing cases that are very different so as to compare using maximum variation. Example: survey only tall and short people
- 5) **Critical case:** important cases that may be repeated elsewhere in the future. Example: catastrophes, or failed students

Summary:

- **Your research question determines the choice of sampling technique.**
- **If you want to generalize to the whole population you will need to use probability sampling, if not non-probability sampling.**
- **The higher the confidence and lower the margin of error you desire, the greater the sample size.**
- **Time, travel and resources are also an important factor – reflecting the importance of access.**

References

- **Alimukhamedova, N. (2022). Research Methods & Data Analysis. MSc Research Methods. Lecture materials. Westminster International University in Tashkent.**
- **Bryman, A. and Bell, E. (2007) Business Research Methods. (3rd edition) Oxford University Press.**
- **Blaxter, L., Hughes, Ch. and Tight, M. (2001). How to research. Maidenhead: OUP.**
- **Kumar, R. (2012). *Research methodology: A step-by-step guide for beginners*. SAGE.**
- **Saunders, M., Lewis, P. and Thornhill, A. (2016). *Research methods for business students*, (7th edition) London: Pearson Education.**

Let's do practice !

Team 1: RQ: What is impact of Covid on household wellbeing ?

Team 2: RQ: Why sales in grocery stores increases suddenly on Fridays ?

Team 3: RQ: What is effect of expensive odor (Chanel №5) on shopping behavior ?

Team 4: RQ: How parents influence on career choice of their children ?

For your group research projects define:

- 1) What is population ?**
- 2) What sample size you would need ?**
- 3) Which sampling technique you would choose ?**