MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Unit 1: Sampling Methods

Mamatha.H.R

Department of Computer Science and Engineering

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Unit 1:Sampling Methods

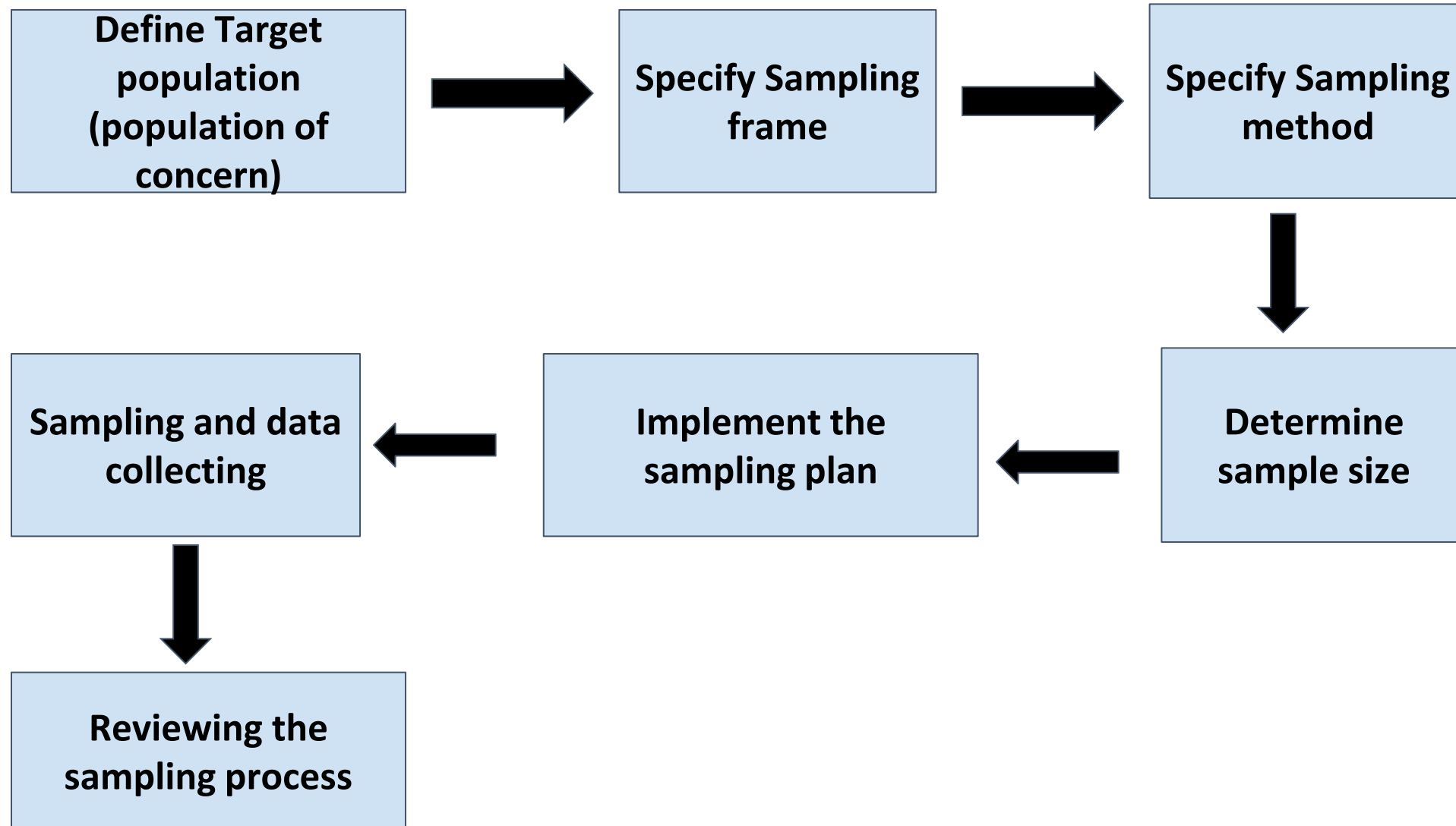**Mamatha H R**

Department of Computer Science and Engineering

Topics to be covered

❖ Sampling methods

❖ Sampling process

❖ Probability and Non-probability sampling

❖ Advantages and disadvantages of different sampling methods

Sampling process

Sampling

➔ **Factors that influence sample representativeness:**
- Sampling procedure
- Sample size
- Participation (response)

➔ **When might you sample the entire population?**
- When your population is very small
- When you have extensive resources
- When you don't expect a very high response

Source: thumbs.dreamstime.com

Sampling Frame

Sampling frame is the <u>list of items or events</u> from which the <u>potential respondents are drawn</u> or which are <u>possible to measure</u>.

- Sometimes, it is possible to identify and measure every single item in the population and to include any one of them in our sample.
- However, in the more general case this is not possible.
- There is no way to identify all rats in the set of all rats.
- As a remedy, we seek a **sampling frame** which has the <u>property that we can identify every single element and include any of them in our sample</u>.
- The sampling frame **must be representative of the population.**

Representative & Biased Sample

Sample 1

Representative of the population
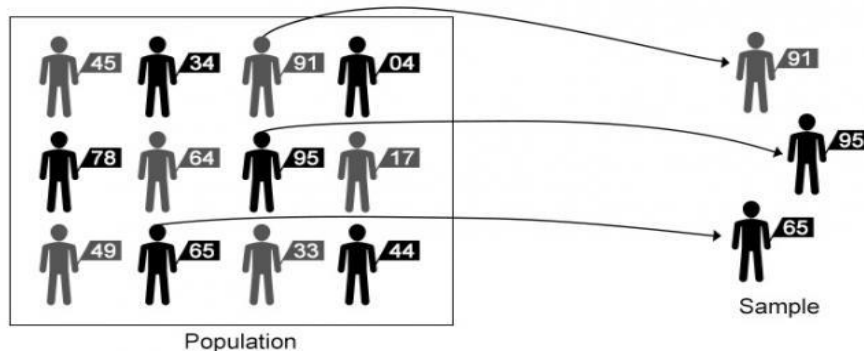
Population

Sample 2

Biased Sample
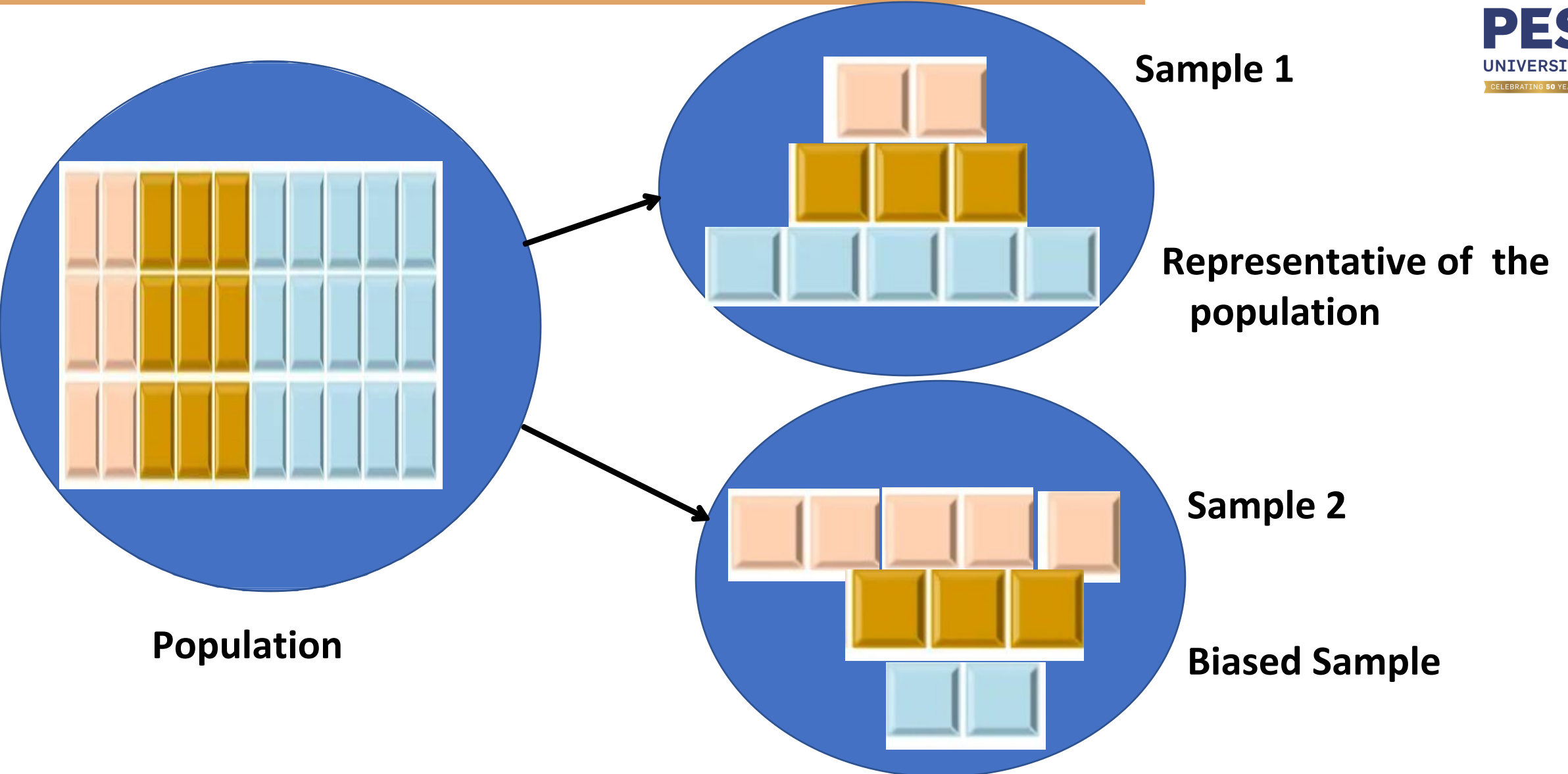
Types of Sampling methods
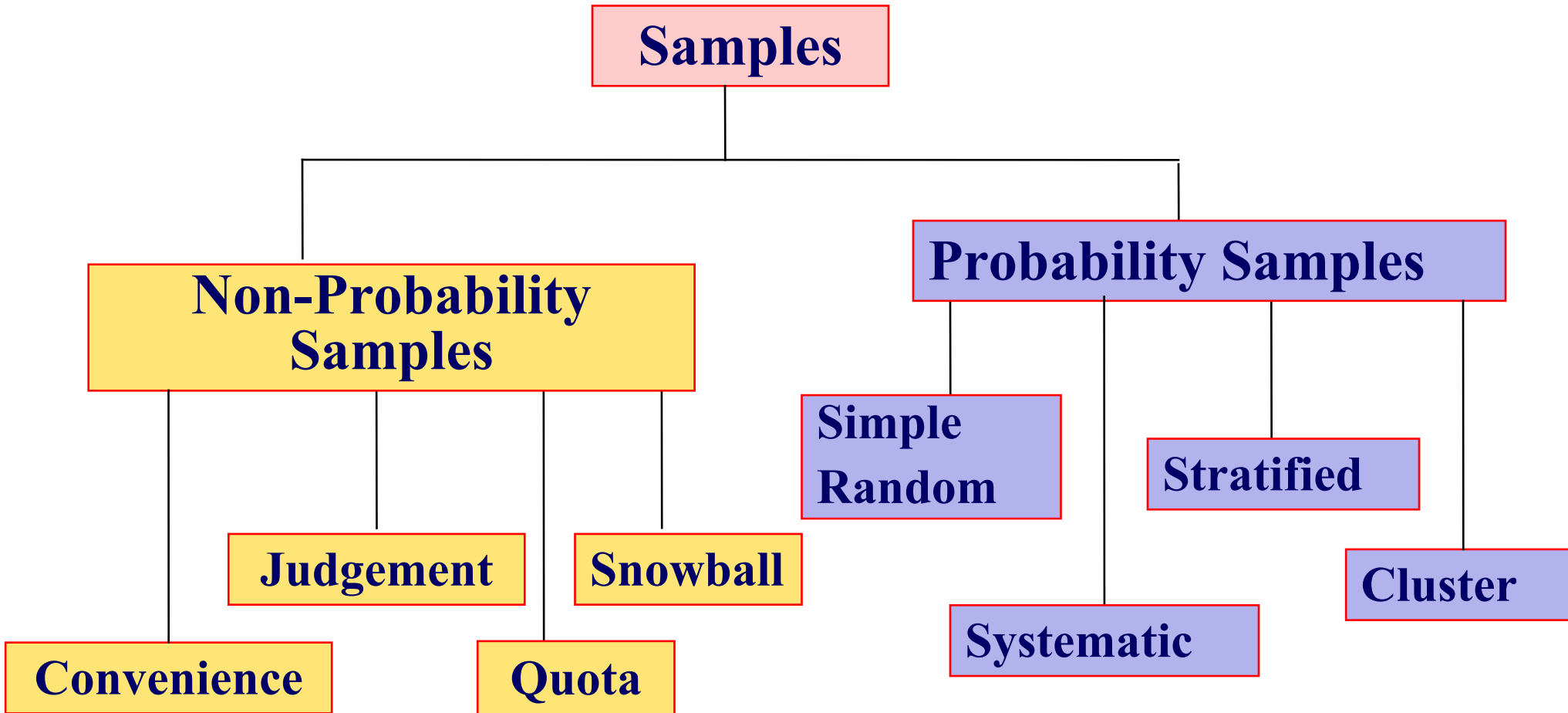
Probability Sampling

- **Probability sampling** is a type of sampling in which <u>every unit in the population has a chance/probability (greater than zero) of being selected in the sample</u>, and this probability can be accurately determined.
- This type of sampling decreases bias and sampling error in the selection process.
- When <u>every element</u> in the population does have the <u>same probability of selection</u>, this is known as an **'equal probability of selection' (EPS)** design. Such designs are also referred to as **'self-weighting'** because <u>all sampled units are given the same weight</u>.

Probability Sampling

Slide Courtesy:Dr.Uma

Non-Probability Sampling

- **Non-Probability sampling** is a type of sampling in which <u>every unit in the population doesn't have a chance/probability (greater than zero) of being selected in the sample</u>.
- Here, <u>some elements of the population have no chance of selection</u> (these are sometimes referred to as **'out of coverage'/'undercovered'**), or the <u>probability of selection can't be accurately determined</u>.
- It involves the **selection** of elements **based on assumptions regarding the population of interest**, which forms the criteria for selection.
- The **selection** of elements is **non random**.
- Thus, non-probability sampling <u>does not allow the estimation of sampling errors</u>.
- It is more likely to produce a **biased sample** and **restricts generalization.**
- It is not an appropriate data collection method for most of the statistical analysis.

Simple Random Sampling

Simple random sampling, as the name suggests, is an <u>entirely random method of selecting the sample</u>.

- Here, each subject or unit in the population has an equal chance of being selected.
- The sampling frame should include the whole population.
- A table of random number or lottery system is used to determine which units are to be selected.
- Simple random sampling is always an EPS design, but not all EPS designs are simple random sampling.

Source: datasciencemadesimple.com
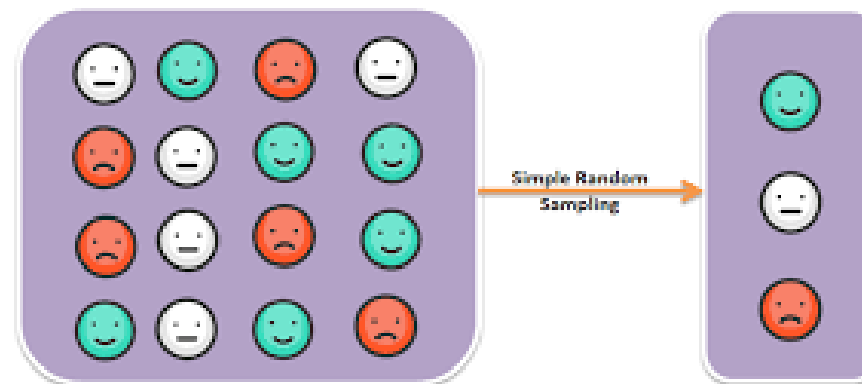
Simple Random Sampling

- **Purpose:** It is random and thus results in a representative-sample.
- **When to Use:** Best to <u>use when population is small</u>
  as it produces a better representative-sample.
- **Key Aspect:** Each member of the population has an equal probability of getting selected.
- **General Procedure:** Assign numbers to all members of the population & select randomly.
  - For a small population: <u>Manual lottery method</u> can be used for selection.
  - For a larger population : <u>System generated numbers</u> can be used to select elements from the population.

Simple Random Sampling Examples

In a company consisting of 10,000 employees, 25 employees are selected to survey the average number of hours a day they are present in the office.

- Population frame: List of all employees numbered from 1-10,000
- Sample : Random number table consisting of 25 random employees.
- Probability of selection of each employee :

N = 10,000; n = 25

probability = (25/10,000) x 100 = 0.25%

Row #	A	B	C	D	E
1	6315	9556	5014	2126	8472
2	434	854	958	2572	9220
3	1961	2395	6882	7590	5329
4	18	8580	9313	454	3511
5	7912	7194	855	753	9016

Simple Random Sampling: Advantages

→ **Advantages:**

● This method is simple to use.

● Estimates are easy to calculate.

● Random samples are usually fairly representative since they don't favor certain members of the population.

● Low sampling error.

● It needs only a minimum knowledge of the study group of population in advance.

Simple Random Sampling: Disadvantages

➔ **Disadvantages:**
- If sampling frame is large, this method impracticable.
- Minority subgroups of interest in population may not be present in sample in sufficient numbers for study.
- This type of sampling can't be employed where the units of the population are heterogeneous in nature.
- Sometimes, it is difficult to have a completely cataloged universe.
- This method lacks the use of available knowledge concerning the population.

Simple Random Sampling with replacement

- This is a sampling procedure in which <u>each sampling unit</u> randomly selected from the population is <u>measured or recorded</u> and then <u>returned to the population</u>. Thus, a sampling unit may be sampled multiple times.
- When sampling the first marble, each marble has the same chance of 0.1 of being sampled. When sampling the second marble and all the subsequent marbles, each marble still has a 0.1 chance of being sampled.
- **Each time we sample a unit, all units have similar chances of being sampled**.

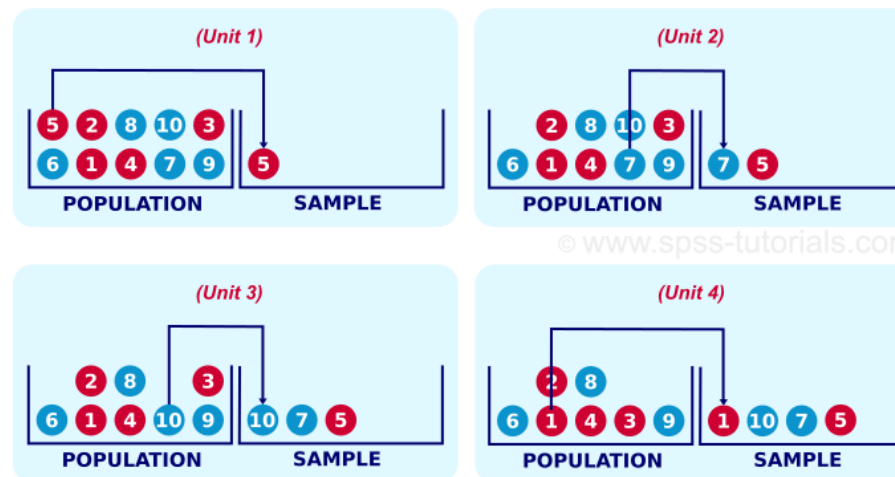SIMPLE RANDOM SAMPLING *WITH* REPLACEMENT
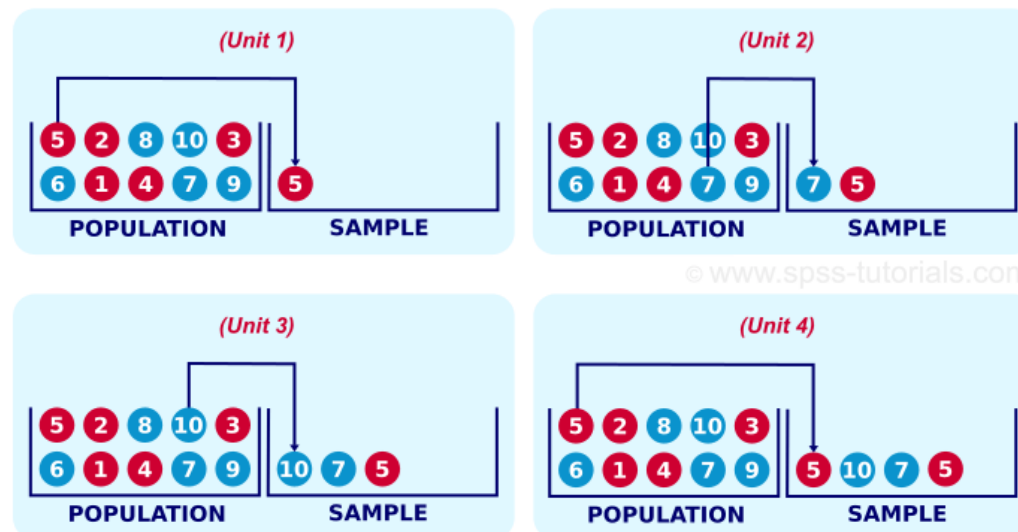
Simple Random Sampling without replacement

- This is a sampling procedure in which sampling units are selected from a population of without replacement such that every sample unit has an equal probability of being selected.
- No element can be selected more than once in the same sample.
- For the first marble sampled, each marble has a 0.1 chance of being sampled. However, the first unit we sampled has a zero chance of being sampled again.
- Thus, the other 9 units each have a chance of 1 in 9 = 0.11 of being sampled as the second unit.

SIMPLE RANDOM SAMPLING *WITHOUT* REPLACEMENT

(Unit 1)

POPULATION   SAMPLE

(Unit 2)

POPULATION   SAMPLE

© www.spss-tutorials.com

(Unit 3)

POPULATION   SAMPLE

*(Unit 4)*

POPULATION   SAMPLE

Systematic Sampling

- Systematic sampling relies on **arranging the target population according to some ordering scheme** and then **selecting elements at regular intervals** through that ordered list.
- The **first element** is selected **randomly**.
- Then it proceeds with the **selection of every kth element.** Where k is the size of the selection interval. $\boxed{\text{k} = \text{(population size/sample size)}}$

- It is important that the starting point is not automatically the first in the list, but is instead randomly chosen from within the first to the kth element in the list.
- A simple example would be to select every 10th name from the telephone directory (an 'every 10th' sample, also referred to as 'sampling with a skip of 10').

Systematic Sampling

- Systematic sampling is an Equal Probability Sampling method, as all elements have the same probability of selection (in the below example given, one in twelve).
- It is not 'simple random sampling' because different subsets of the same size have different selection probabilities
- Ex: the set {2,5,8,11} has a one-in-twelve probability of selection, but the set {1,3,6,7} has zero probability of selection.

Population

Sample (every 3rd)

Source: www.netquest.com
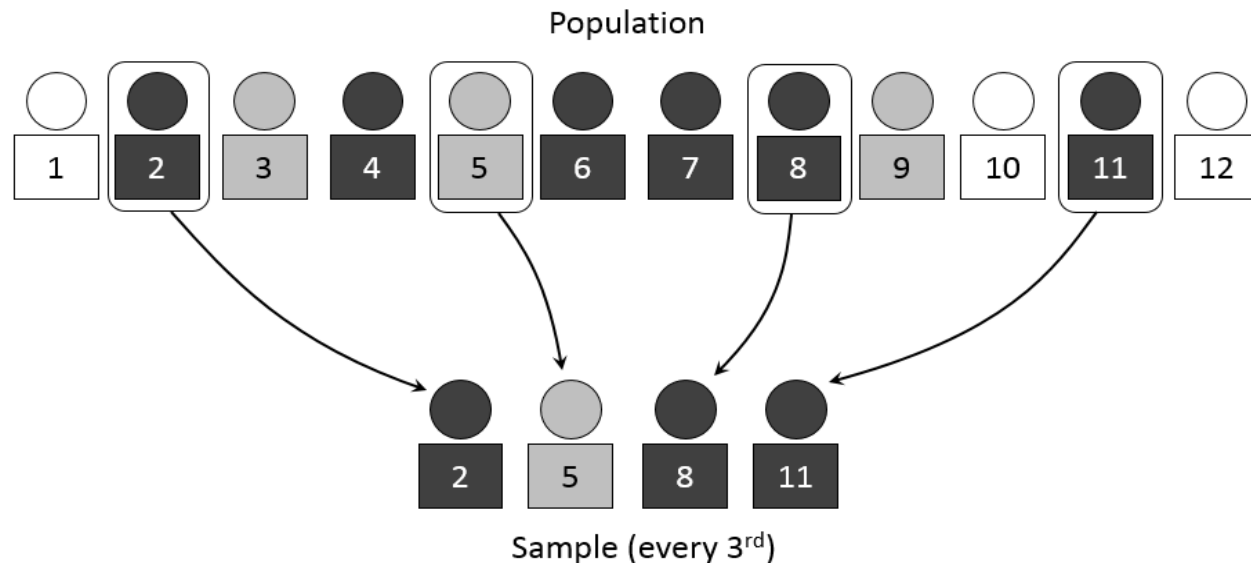
Systematic Sampling

- **When to Use:** When project budget is tight and less time to complete.
- **Key Aspect:** Find the kth value to select every kth member.

  k = N / n
- **General Procedure:**
  - Assign numbers to each population element.
  - Order the population elements in an ordered sequence
  - Find 'k' the size of the selection interval.
  - Select the first sample element randomly from the first k population elements.
  - Thereafter, select the sample elements at a constant interval, k, from the ordered sequence frame.

Systematic Sampling: Advantages

- Sample is easy to select.
- Suitable sampling frame can be identified easily.
- Sample evenly spreads over entire reference population.
- It is a cost effective sampling method.
- It guarantees that the entire population is evenly sampled.
- Systematic sampling also carries a low-risk factor because there is a low chance that the data can be contaminated.

Systematic Sampling: Disadvantages

- This type of sampling **might lead to bias** if there is an underlying pattern/periodicity in the population which coincides with the selection.

  Ex : If the HR database groups employees by team, and team members are listed in order of seniority, there is a risk that the interval might skip over people in junior roles, resulting in a sample that is skewed towards senior employees.

- Difficult to assess precision of estimate from one survey.
- Each element does not have an equal chance in getting selected
- Ignorance of all the elements between two kth elements.
- The size of the population is needed. Without knowing the specific number of participants in a population, systematic sampling does not work well.

Stratified Sampling

● Stratified sampling is the type of sampling in which the population is divided into 2 or more groups called strata based on a shared characteristic or trait.

● Then simple random samples are selected from each group.

● The selected 2 or more samples are combined into one.

● The strata or groups don't overlap. But, they represent the entire population.

● The shared characteristics based on which the population is divided could be gender, educational attainment, income, age etc.

Source: datasciencemadesimple.com
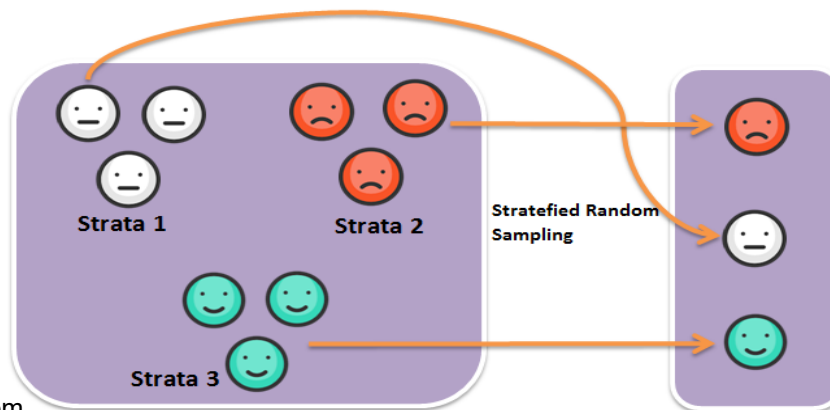
Stratified Sampling

- **Each stratum** is sampled as an **independent sub-population.**
- Every unit in a stratum has same chance of being selected.
- Using same sampling fraction for all strata **ensures proportionate representation in the sample**.
- Adequate representation of minority subgroups of interest can be ensured by stratification & varying sampling fraction between strata as required.
- Since each stratum is treated as an independent population, different sampling approaches can be applied to different strata.
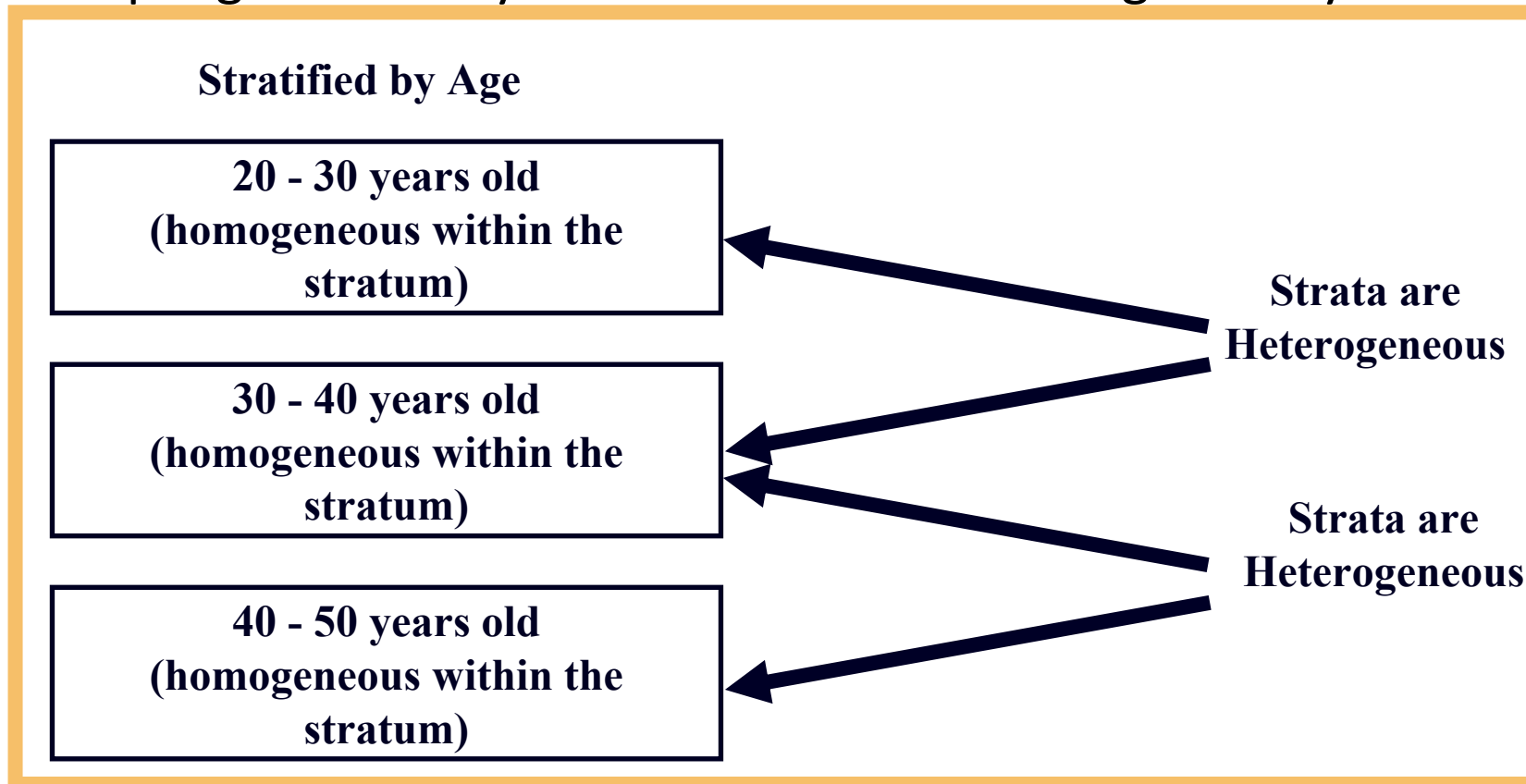
Stratified Sampling

- **Purpose**: To obtain an unbiased random sample from a larger population.
- **When to Use**: When population proportion must be reflected in sample.
- **Key Aspect**: Sample proportion is same as Population proportion, Strata is homogeneous.
- **General Procedure:**
  - Divide the population into Strata or Groups.
  - Criteria for division could be: Gender, Hair Color, Eye Color, Salary, Designation, Age etc.
  - Selection of sample: Simple Random Sampling approach is used to sample units from each strata.

Stratified Sampling examples

To find out the most popular song among the FM radio listeners.

- All listeners are stratified by age.
- Listeners from each age group are selected using simple random sampling and surveyed for their favourite song of the year.

Stratified by Age

| 20 - 30 years old (homogeneous within the stratum) |
| 30 - 40 years old (homogeneous within the stratum) |
| 40 - 50 years old (homogeneous within the stratum) |

Strata are Heterogeneous

Strata are Heterogeneous

Stratified Sampling: Advantages

- It enhances the representativeness of the sample.
- It is easy to carry out.
- It has higher statistical efficiency.
- A stratified sample can provide a higher precision than a simple random sample of the same size.
- As it provides a greater precision, this type of sampling often requires a smaller sized sample which saves money.
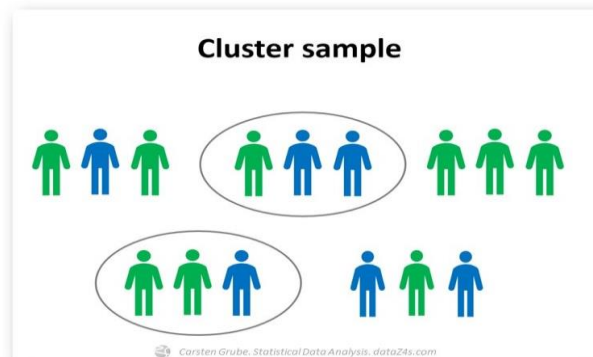
Stratified Sampling: Disadvantages

- Sampling frame of the entire population has to be prepared separately for each stratum.
- When examining multiple criteria to divide the population, stratifying variables may be related to some but not to others further complicating the design and potentially reducing the utility of the strata.
- In some cases (such as designs with a large number of strata, or those with a specified minimum sample size per group), stratified sampling can potentially require a larger sample than other methods.
- It is time consuming and expensive.
- It leads to classification errors.

Cluster Sampling

- In cluster sampling, population is divided into non-overlapping clusters or areas similar to Stratified sampling.
- Each cluster is a miniature or microcosm of the population.
- Each cluster should have similar characteristics to the whole sample.
- Instead of sampling individuals from each subgroup like in stratified sampling, in cluster sampling entire clusters are randomly selected.
- A subset of the clusters is selected randomly for the sample.
- If the number of elements in the subset of clusters is larger than the desired value of n(sample size), these clusters may be subdivided to form a new set of clusters and subjected to a random selection process.

Cluster sample

Carsten Grube. Statistical Data Analysis. dataZ4s.com

Source: dataz4s.com

Slide Courtesy:Dr.Uma

Cluster Sampling

- **When to Use**: When population is already broken up into groups(clusters).
- **Key Aspect:** Heterogeneous members in each group.
- **General Procedure:**
  - Population is divided into non-overlapping areas(clusters).
  - Each cluster is a miniature or microcosm of a population.
  - Clusters are selected randomly.
  - All elements of the selected-clusters are included in the sample or elements from the selected-clusters are chosen using simple random sampling.

Cluster Sampling examples

The municipal council of a small city wants to investigate the use of health care services by residents.

- The council first obtains electoral subdivision maps that identify and label each city block. From these maps, the council creates a list of all city blocks. This list will serve as the sampling frame.

- Every household in that city belongs to a city block, and each city block represents a cluster of households. The council randomly picks a number of city blocks.
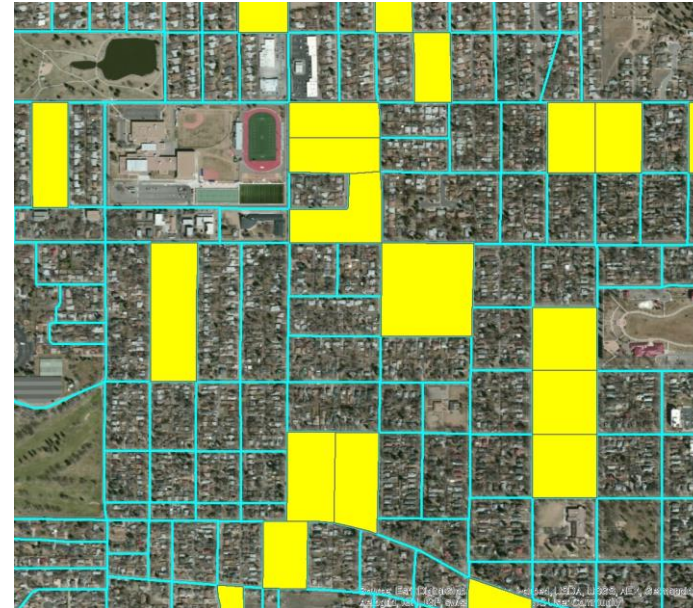
Cluster Sampling: Advantages

- It is more convenient for geographically dispersed populations.
- It can **reduce the travel costs** to contact sample elements.
- It **simplifies the administration of the survey**.
- It is **more feasible**. The division of the entire population into homogeneous groups increases the feasibility of the sampling.
- Since each cluster represents the entire population, **more subjects can be included in the study**.
- **Requires fewer resources**. Since cluster sampling selects only certain groups from the entire population, the method requires fewer resources for the sampling process.

Cluster Sampling: Disadvantages

- It is statistically less efficient when the cluster elements are similar.
- Costs and the number of problems occurring are greater than that of simple random sampling.
- There is higher sampling error.
- The method is **prone to biases**. If the clusters representing the entire population were formed under a biased opinion, the inferences about the entire population would be biased as well.
- It's difficult to guarantee that the sampled clusters are really representative of the whole population.

Non-probability Sampling

Non-Probability sampling is a type of sampling in which every unit in the population doesn't have a chance/probability (greater than zero) of being selected in the sample.

Non-Probability Samples

Convenience

Judgement

Quota

Snowball
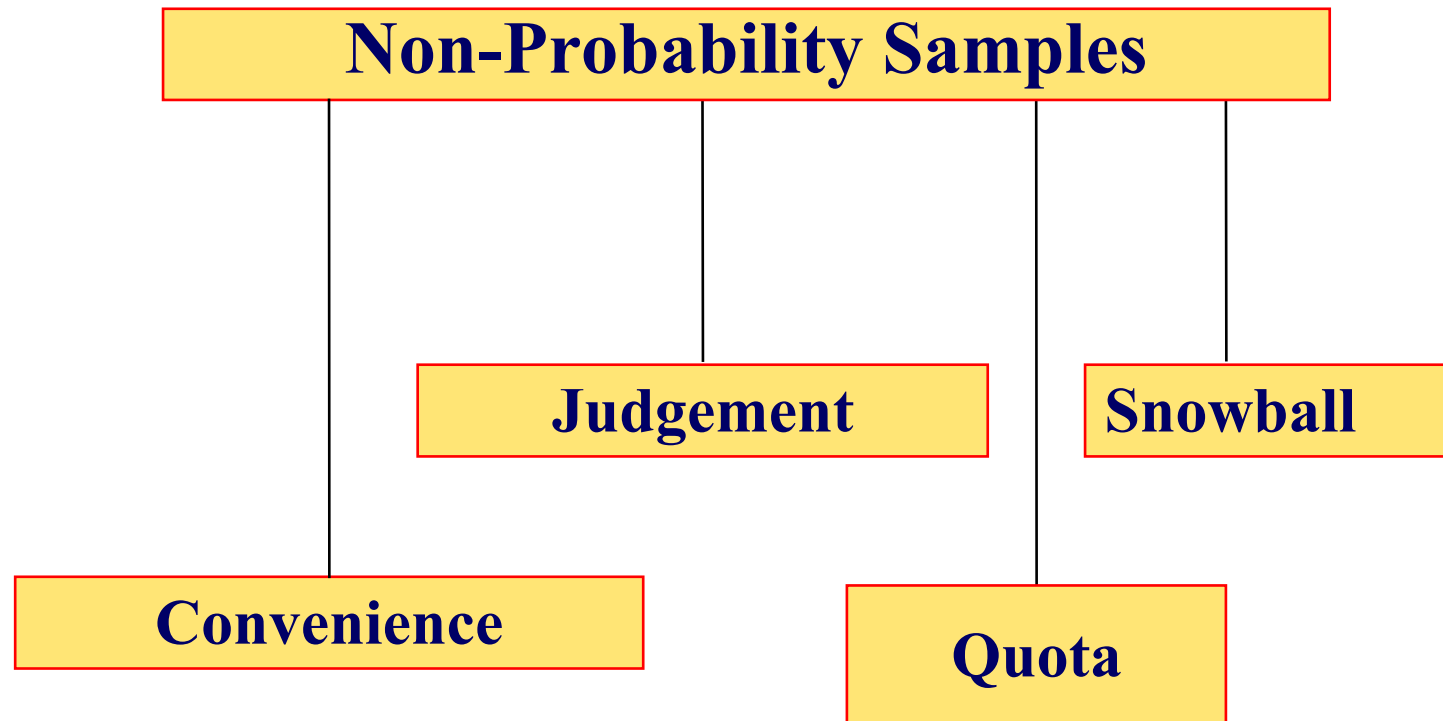
Convenience Sampling

- Sometimes it is also known as **grab or opportunity sampling** or **accidental or haphazard sampling**.
- This is a type of nonprobability sampling which involves the sample being drawn from that part of the population which is close to hand. That is, readily available and convenient.
- Here, sample elements are selected for the convenience of the researcher.
- The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because it would not be representative enough.

Convenience sampling

Source: googleusercontent.com
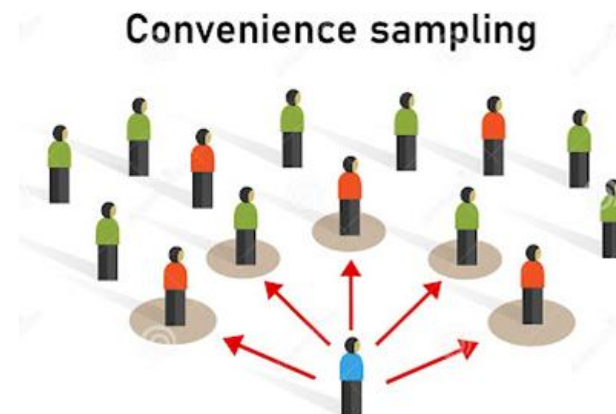
Convenience Sampling

- **When to Use:** When population is not clearly defined or sampling unit is not clear or complete source list is not available.
- **Key Aspect:** Subjects for a study are easily available within the proximity of the researcher.
- **General procedure:**
  - It is done at the "convenience" of the researcher.
  - Selection : The individuals that are convenient and easiest to reach are selected to be included in the sample.

Convenience Sampling examples

To record the popular opinions of people about the current laws of the city.

- The researcher surveys all people that pass by his house.
- Again, this is a convenient way of studying the opinions of people living in the city. But, it doesn't reflect the opinions of all the residents of the city.

Source:slideshare.com
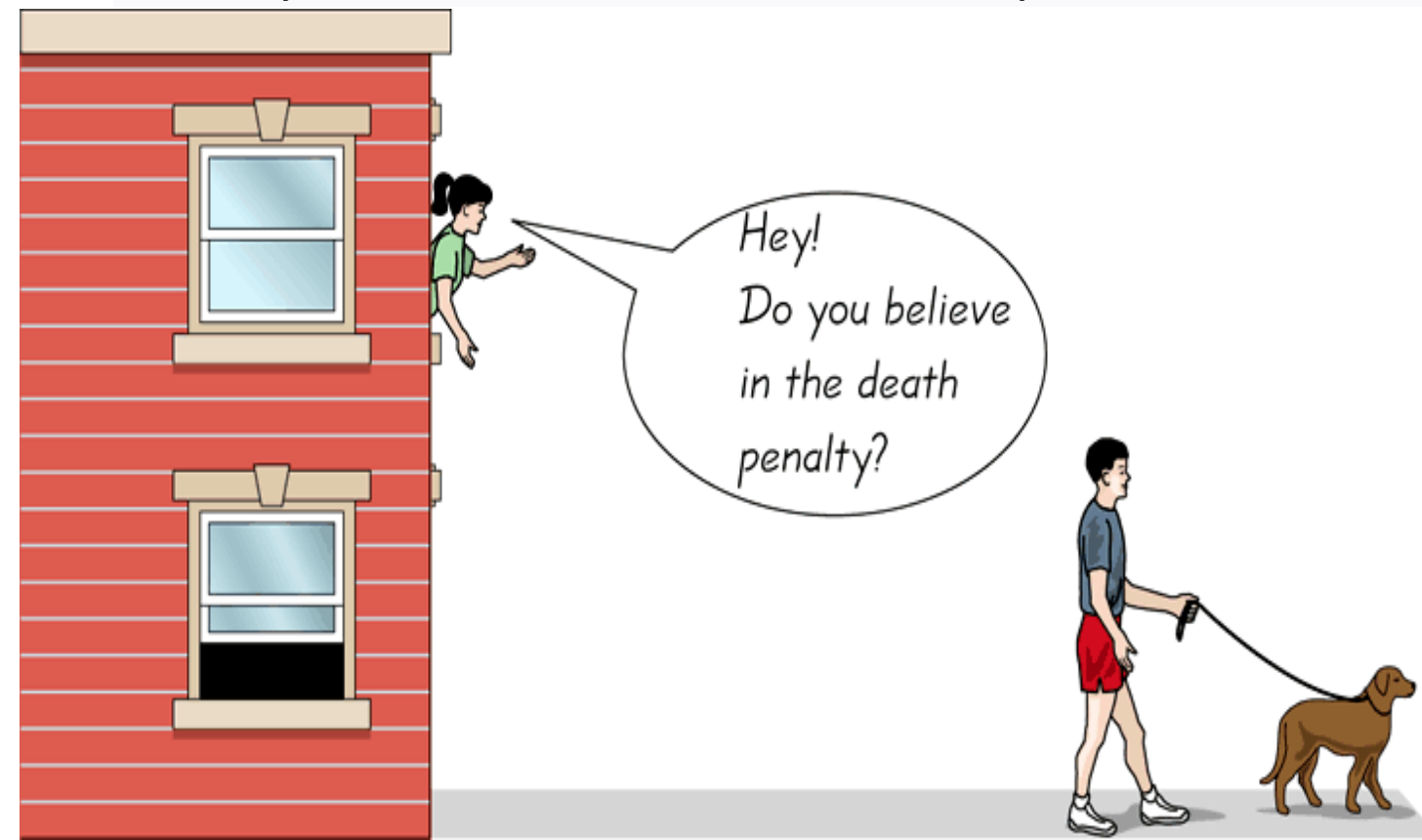
## Convenience Sampling: Advantages & Disadvantages

➔ **Advantages:**
- This type of sampling is useful in pilot study.
- It costs less and is an inexpensive way to gather initial data for the research.
- It saves time.
- It is relatively easy to get a sample.
- It is simple and easy to implement.

➔ **Disadvantages:**
- It is prone to significant bias as the sample may not be representative of the characteristics of the population.
- Since the same may not be representative of the population, this type of sampling can't produce generalizable results.
- It might lead to sampling errors.
- A study conducted on a convenience sample will have limited external validity.

Judgemental Sampling

- Judgemental or Purposive sampling is a type of non-probability sampling where the **researcher chooses the sample** based on who they think would be appropriate for the study.
- This is used primarily when there is a limited number of people that have expertise in the area being researched.
- The sample depends on the judgement of the experts conducting the study.
- It is not a scientific method of sampling.

Judgement sample

Carsten Grube. Statistical Data Analysis. dataZ4s.com
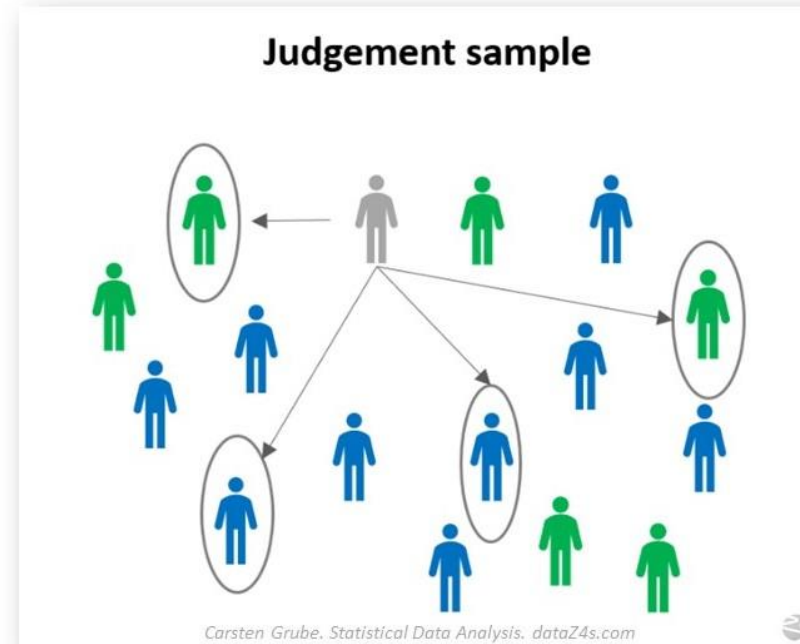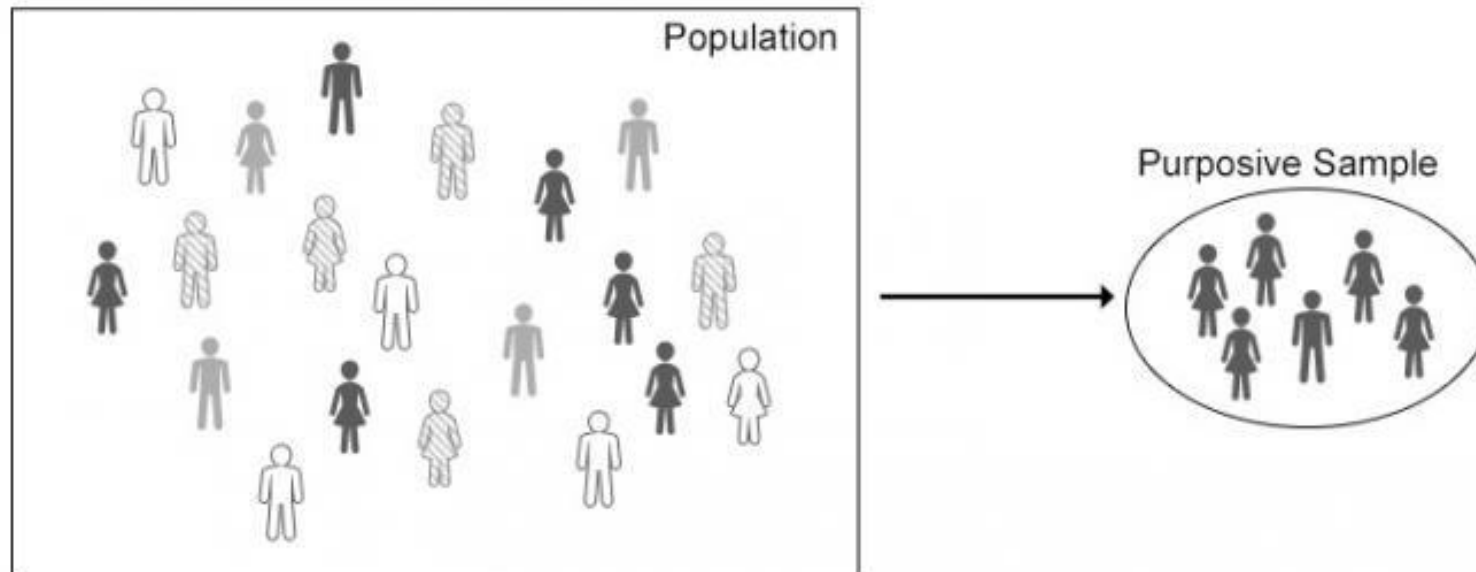
Judgemental Sampling

- **When to Use:** This is used primarily when there is a limited number of people that have expertise in the area being researched.
  Also, the researcher must be confident that the chosen sample is truly representative of the entire population.
- **Key Aspect:** The researcher selects a sample based on experience or knowledge of the group to be sampled.
- **General Procedure:**
  - On the basis of the researcher's knowledge and judgment elements of the population are sampled.
  - Selection : Elements that own the qualities expected by the researcher.

Judgemental Sampling examples

To know more about the opinions and experiences of disabled students at your university

- You purposefully select a number of students with different support needs at your university in order to gather a varied range of data on their experiences with student services.

Purposive Sampling

Judgemental Sampling: Advantages & Disadvantages

➔ **Advantages:**
- It consumes minimum time.
- The researcher is given an opportunity to bring his judgement and expertise to play.
- No special knowledge of statistics is needed.
- Real time results can be obtained.

➔ **Disadvantages:**
- It is prone to errors in judgment by researcher.
- Low level of reliability and high levels of bias.
- Inability to generalize research findings to the entire population.
- It is difficult to choose the appropriate sample size.

Quota sampling

- In this type of sampling, sample elements are <u>selected until the quota controls are satisfied.</u>
- The population is first segmented into mutually exclusive sub-groups, just as in stratified sampling.
- Then judgment is used to select subjects or units from each segment based on a specified proportion.
- The population units are selected based on predetermined characteristics of the population.
- It is similar to Stratified sampling but it doesn't involve random selection.
- Ex: recruiting the first 50 men and first 50 women that meet inclusion criteria.
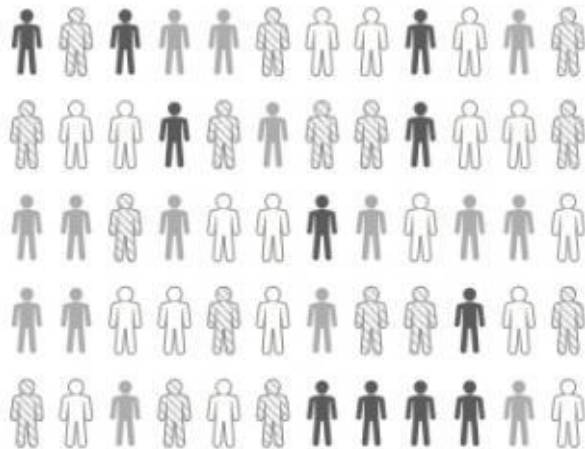
Quota sampling

- **When to Use:** If a study aims to investigate a trait or a characteristic of a certain subgroup, this type of sampling is the ideal technique.
- **Key Aspect:** Sample elements are selected until the quota controls are satisfied.
- **General Procedure:**
  - Divide the population into subgroups.
  - Identify proportions or weightage in which the subgroups are present in the population.
  - Select an appropriate sample size while maintaining the proportions of the subgroups.
  - Conduct the surveys according to the quotas defined

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Quota sampling examples

A cool drinks company wants to find out what age group prefers what brand of drinks in a particular city.

- The researcher applies quotas on the age groups of 11-21,22-31, 32-41, 42-51.
- The researcher then samples people from each quota and surveys them to gauge the trend among the population of the city.

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

Quota sampling examples: Advantages & Disadvantages

➔ **Advantages:**
- It is a cost effective method.
- There is convenience in execution of this sampling.
- It is a speedy process.
- The information can be deciphered once the sampling is done.
- It improves the representation of certain groups within the population and also ensures that they are not over-represented.

➔ **Disadvantages:**
- Impossible to determine sampling error as the sample is not chosen using random selection.
- Can result in sampling bias if the selection of units was based on ease of access and cost considerations.
- It is not possible to make statistical inferences from the sample to the population leading to the problems of generalization.

Slide Courtesy:Dr.Uma

Snowball sampling

- In this type of sampling, survey subjects are selected based on referral from other survey respondents.
- Existing subjects are asked to nominate further subjects known to them so that the sample increases in size like a rolling snowball.
- This method of sampling is effective when a sampling frame is difficult to identify.
- Usually applied when the subjects are difficult to trace. Ex: it will be extremely challenging to survey shelter less people or illegal immigrants.

Snowball sampling

Source: cuttingedgepr.com
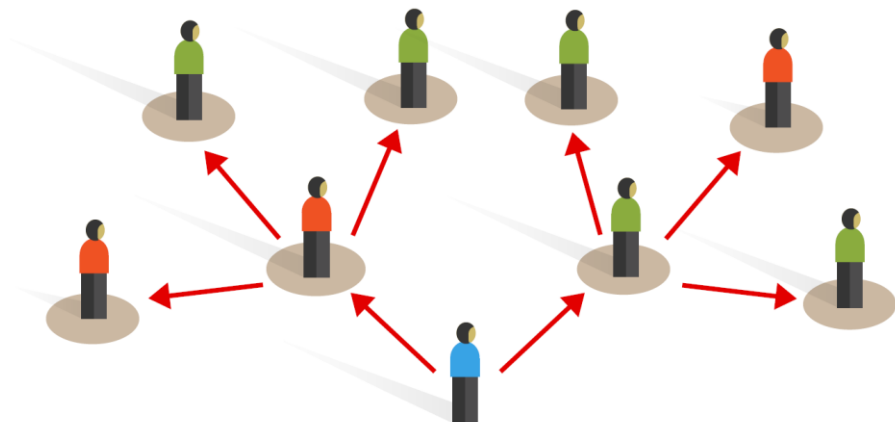
Snowball sampling

- **When to Use:** When the desired sample characteristic is rare.
- **Key Aspect:** Research starts with a key person and introduce the next one to become a chain. It may be extremely difficult or cost prohibitive to locate respondents in these situations.
- **How:**
  - Identify an initial subject and ask these people to identify others.
  - Selection : This technique relies on referrals from initial subjects to generate additional subjects.

Snowball sampling examples

To study the level of customer satisfaction among the members of an elite   country club.

- It is extremely difficult to collect primary data sources unless a member of the club agrees to have a direct conversation with you and provides the contact details of the other members of the club.
- Thus the primary data source is randomly selected and it nominates other potential data sources that will be able to participate in the research studies.
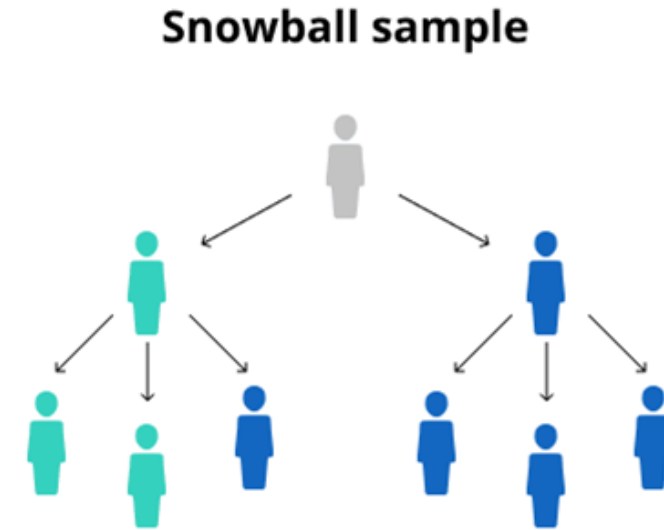
Snowball sample

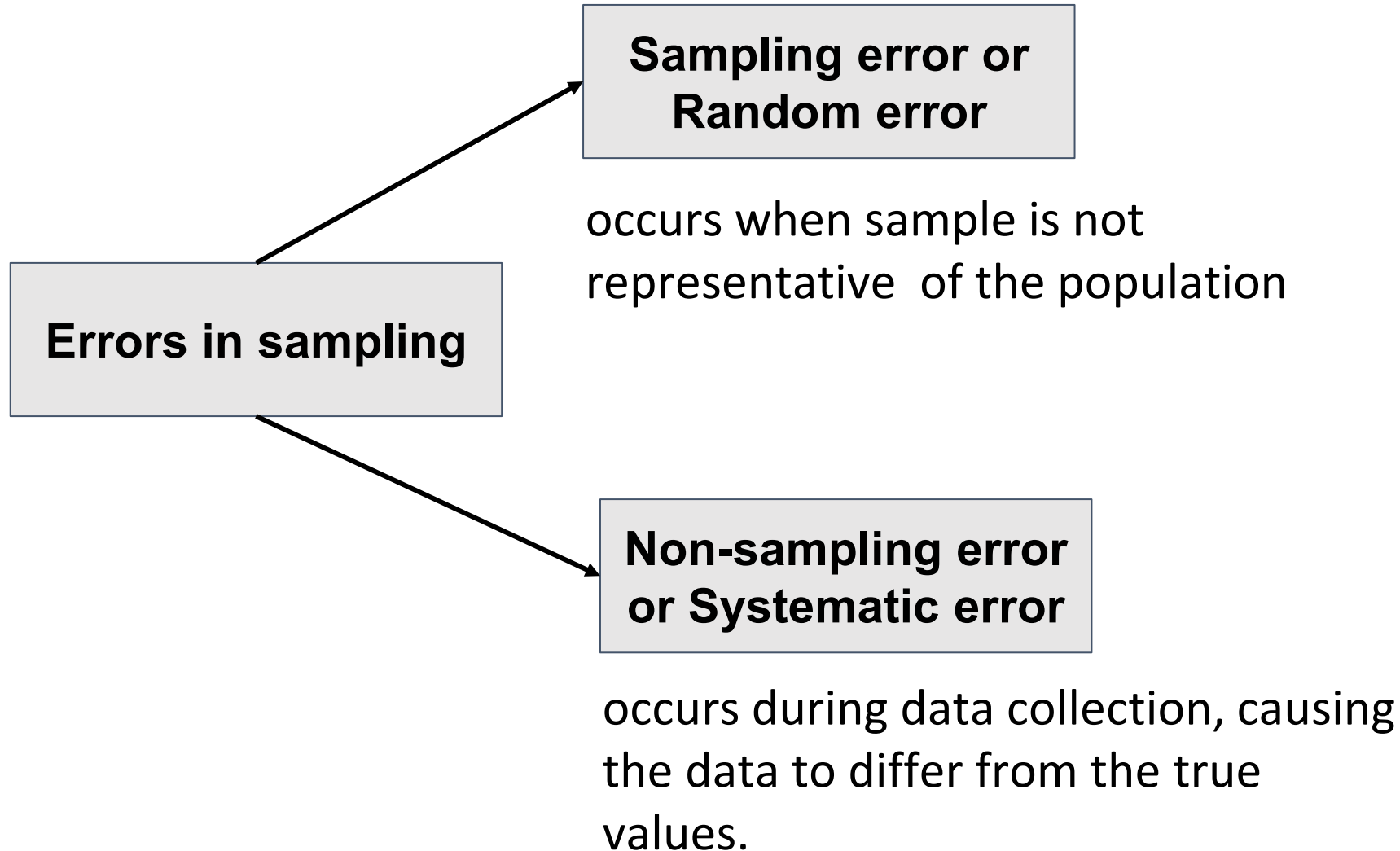## Snowball sampling examples: Advantages & Disadvantages

➔ **Advantages:**
- The chain referral process allows the researcher to reach populations that are difficult to sample when using other sampling methods.
- The process is cheap, simple and cost-efficient.
- This sampling technique needs little planning and fewer workforce compared to other sampling techniques.

➔ **Disadvantages:**
- There is a significant risk of selection bias in snowball sampling, as the referenced individuals will share common traits with the person who recommends them.
- It is usually impossible to determine the sampling error or make inferences about populations based on the obtained sample.
- The researcher has little control over the sampling method.
- Representativeness of the sample is not guaranteed.

Errors in sampling

Sampling error or Random error

Errors in sampling

occurs when sample is not representative of the population

Non-sampling error or Systematic error

occurs during data collection, causing the data to differ from the true values.

Slide Courtesy:Dr.Uma

Sampling error

- The discrepancy between a sample statistic and its population parameter is called sampling error.
- Defining and measuring sampling error is a large part of inferential statistics.
- It occurs when the sample is not representative of the population.
- The sampling error for a given sample is unknown but when the sampling is random, for some estimates (for example, sample mean, sample proportion) theoretical methods may be used to measure the extent of the variation caused by sampling error.

Sampling error

Population
of 1000 college students

Population Parameters
Average Age = 21.3 years
Average IQ = 112.5
65% Female, 35% Male

Sample #1

Eric
Jessica
Laura
Karen
Brian

Sample Statistics
Average Age = 19.8
Average IQ = 104.6
60% Female, 40% Male

Sample #2

Tom
Kristen
Sara
Andrew
John

Sample Statistics
Average Age = 20.4
Average IQ = 114.2
40% Female, 60% Male
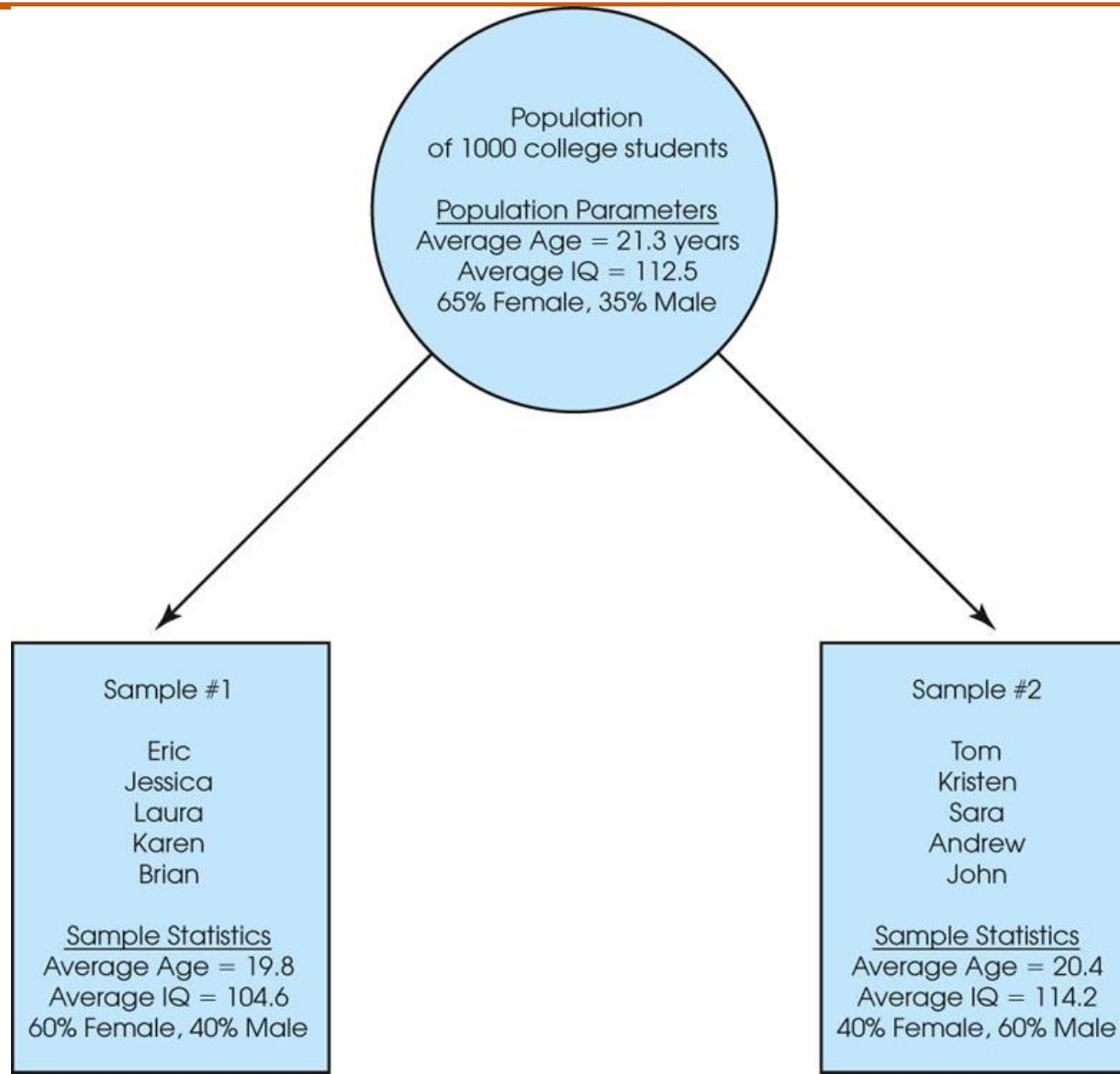
As we can see there is a difference between population parameters and sample parameters. This is due to sampling error.

Two samples of same population have differing parameters. This is due to sampling variation. It is also the reason why scientific experiments produce different result under identical scenarios.

Non-sampling error

- Non-sampling errors are the results of mistakes made in implementing data collection and data processing, such as
  - failure to locate and interview the correct household
  - errors in understanding of the questions by either the interviewer or the respondent
  - data entry errors
  - missing Data
  - poorly conceived concepts, unclear definitions, and defective questionnaires
  - response errors occurring when people are unaware, refuse to answer, or overstate in their answers
- Major sources : Sampling Bias, Non-response Bias.

Sampling Bias

- Sampling bias occurs when a chosen sample is not representative of the larger population.
- It occurs due to the sampling technique/method used to perform data collection.
- It can be either **selection bias and non-response bias.**
- A sampling method has a sampling bias if all subjects in the population are not equally likely to be included in a sample.
- That is, a sample is collected in such a way that some members of the intended population have a lower or higher sampling probability than others.

Selection Bias & Nonresponse bias

➔ **Selection bias:**
- It is a bias in which a sample is collected in such a way that some members of the intended population have a lower or higher sampling probability than others.
- It results in a biased sample of a population in which all individuals, or instances, were not equally likely to have been selected.

➔ **Nonresponse bias:**
- Nonresponse bias is a type of sampling bias that occurs because of the absence of certain objects or subjects from a sample.
- For example, some subjects don't respond to surveys because they refuse, cannot be contacted, or have a lack of interest in the survey content.

Bias ex:

Q) A new chemical process is run 10 times each morning for five consecutive mornings. If the new process is put into production, it will be run 10 hours each day, from 7 A.M. until 5 P.M. Is it reasonable to consider the 50 yields to be a simple random sample?

Ans) Since the new process runs during both morning and afternoon, the population consists of all the yields that would ever be observed, including both morning and afternoon runs.

The sample however is drawn only from that portion of the population that consists of morning runs, and thus it is not a simple random sample. It exhibits a bias is not representative of the population intended to be studied.

Sampling variation

- Simple random samples always differ from their populations in some ways, and occasionally may be substantially different.
- Two different samples from the same population will differ from each other as well.
- This phenomenon is known as **sampling variation**.
- Sampling variation is one of the reasons that scientific experiments produce somewhat different results when repeated, even when the conditions appear to be identical.

Independence

- The items in a sample are said to be independent if knowing the values of some of them does not help to predict the values of the others.
- With a finite, tangible population, the items in a simple random sample are not strictly independent, because as each item is drawn, the population changes.
- This change can be substantial when the population is small.
- However, when the population is very large, this change is negligible and the items can be treated as if they were independent
- The sample can be considered independent if sample size is smaller than 5% of population size.
- Since conceptual population have infinite/very large size the sample obtained (ex: measuring a rock) is always independent

Questions

(Q1.) A physical education professor wants to study the physical fitness levels of students at her university. There are 20,000 students enrolled at the university, and she wants to draw a sample of size 100 to take a physical fitness test. She obtains a list of all 20,000 students, numbered from 1 to 20,000. She uses a computer random number generator to generate 100 random integers between 1 and 20,000 and then invites the 100 students corresponding to those numbers to participate in the study. Which sampling technique is used?

Answer:

The simple random sampling technique is used.

Note that it is analogous to a lottery in which each student has a ticket and 100 tickets are drawn.

Questions

(Q2) A quality engineer wants to inspect rolls of wallpaper in order to obtain information on the rate at which flaws in the printing are occurring. She decides to draw a sample of 50 rolls of wallpaper from a day's production. Each hour for 5 hours, she takes the 10 most recently produced rolls and counts the number of flaws on each. Is this a simple random sample?

Answer:

No. Not every subset of 50 rolls of wallpaper is equally likely to comprise the sample. To construct a simple random sample, the engineer would need to assign a number to each roll produced during the day and then generate random numbers to determine which rolls comprise the sample.

Questions

(Q3) A construction engineer has just received a shipment of 1000 concrete blocks, each weighing approximately 50 pounds. The blocks have been delivered in a large pile. The engineer wishes to investigate the crushing strength of the blocks by measuring the strengths in a sample of 10 blocks. Which sampling method is suitable?

Answer:

To draw a simple random sample would require removing blocks from the center and bottom of the pile, which might be quite difficult. For this reason, the engineer might construct a sample simply by taking 10 blocks off the top of the pile.

convenience sample

Questions

(Q4) A quality inspector draws a simple random sample of 40 bolts from a large shipment and measures the length of each. He finds that 34 of them, or 85%, meet a length specification. He concludes that exactly 85% of the bolts in the shipment meet the specification. The inspector's supervisor concludes that the proportion of good bolts is likely to be close to, but not exactly equal to, 85%. Which conclusion is appropriate?

Answer:

Because of sampling variation, simple random samples don't reflect the population perfectly. However, they are often fairly close. It is therefore appropriate to infer that the proportion of good bolts in the lot is likely to be close to the sample proportion, which is 85%. It is not likely that the population proportion is equal to 85%.

Questions

(Q5) Another inspector repeats the study with a different simple random sample of 40 bolts. She finds that 36 of them, or 90%, are good. The first inspector claims that she must have done something wrong, since his results showed that 85%, not 90%, of bolts are good. Is he right?

Answer:

No, he is not right. This is sampling variation at work. Two different samples from the same population will differ from each other and from the population.

Questions

(Q6) A geologist weighs a rock several times on a sensitive scale. Each time, the scale gives a slightly different reading. Under what conditions can these readings be thought of as a simple random sample? What is the population?

Answer:

If the physical characteristics of the scale remain the same for each weighing, so that the measurements are made under identical conditions, then the readings may be considered to be a simple random sample. The population is conceptual. It consists of all the readings that the scale could in principle produce.

Questions

(Q7) What sampling method can be recommended?

- Determining proportion of undernourished five year olds in a village.
- Investigating nutritional status of preschool children.
- In estimation of immunization coverage in a province, data on seven children aged 12-23 months in 30 clusters are used to determine proportion of fully immunized children in the province.Give reasons why cluster sampling is used in this survey.

References

https://www.spss-tutorials.com/simple-random-sampling-what-is-it/

https://www.analyticsvidhya.com/blog/2019/09/data-scientists-guide-8-types-of-sampling-techniques/

Text Book:

Statistics for Engineers and Scientists, William Navidi.

THANK YOU

Dr.Mamatha H R

Professor, Department of Computer Science

mamathahr@pes.edu

+91 80 2672 1983 Extn 712