بسم الله الرحمن الرحيم

Biostatistics

Lecture five

Sampling methods

Doctor:

Israa rawashdeh

Done by Doctor:

Abdullah daradkh

Mahmoud Otoom

Sampling

Terminology:

An **<u>element</u>** is an object on which a measurement is taken

<u>Target population</u> is the collection of the elements which has some or the other characteristic in common (The complete group of interest). Number of elements in the population is the size of the population.

A **sampling frame** is the group of individuals or objects from which a sample is drawn. It is a list of all <u>units</u> within a population who ***can be*** sampled, and may include individuals, households or institutions
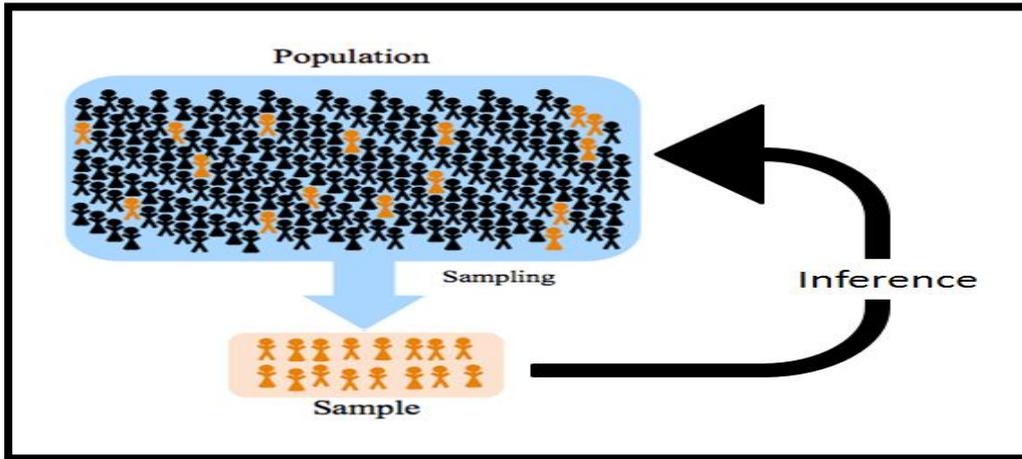
<u>Sample</u> is the subset of the population. Number of elements

the sample is the sample size
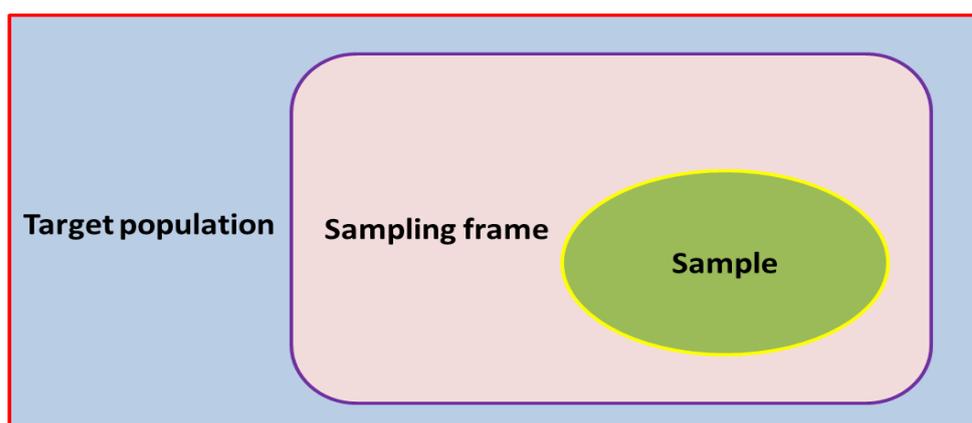
Sampling Methods (طرق أخذ العينات)

Sampling is the process of selecting <u>observations</u> (a sample) to provide an adequate description and inferences of the population.

→ **Sampling** is **a crucial issue** in health research.

Inferential statistics makes inferences about populations using data drawn from the population. Instead of using the entire population to gather the data, the statistician will collect a sample or samples from the millions of residents and make inferences about the entire population using the sample.

- مثلا نقوم بأخذ عينة من الناس و نجري عليهم بحثا و نقوم بتعميم نتائج البحث على مختلف الناس ضمن ذاك المجتمع حتى لو لم يكونوا من ضمن العينة المختارة.

Sample:
- It is a unit that is **selected** from **population**
- **Represents** the whole population
- Purpose to draw **the inference**
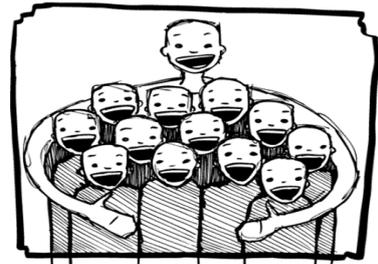
Why do we need to select a sample?

1-The population of interest is usually **too large** to attempt to **survey** all of its members. (we can't take data from too large number of people)

2-A carefully chosen sample can be used to represent the population (Generalize findings on all of the population).

3-Resources (time, money) and **workload.** (because we work on a **small** sample of the population)

4-Gives results with known **accuracy** that can be **calculated mathematically.**

Advantages of sampling are:

In some Cases, the sample could be the whole population as census & rare diseases (we take all of the population because it is difficult to find the disease in a small sample)

1-The cost is lower.

2-The data collection is faster than measuring the entire population.

The sampling process comprises several stages:

1-Defining the population of concern. (population could be people or items)

* A population can be defined as including all elements (people or items) with the characteristic one wishes to understand.

* An element is an object on which a measurement is taken.

2-Specifying a sampling frame. (is the source material or device from which a sample is drawn. It is a list of all units within a population who can be sampled, and may include individuals, households or institutions)

* A sample is a collection of sampling units drawn from a sampling frame.

Example :

If we want to know the average intake of folic acid for pregnant women in Jordan population.

In this case the population is the pregnant Jordanian women

But the sampling frame is the women who take folic acid

- تكون العينة من النساء الحوامل أما اطار العينة يكون من النساء الحوامل الائي يأخذن folic acid

3-Specifying a sampling method for selecting items or events from the frame

4-Determining the sample size

5-Implementing the sampling plan (the way of collection information ,,,on paper, record……..)

6-Sampling and data collection

There are two groups of sampling:

Group 1 : A- Probability .

Group 2 : ,   B- non-probability.

Group 1 : A- Probability احتمالية

A probability sample is a sample in which every unit in the population has a <u>**chance**</u> (greater than zero) of being selected in the sample, and this probability can be accurately determined. (selected randomly).

و **هذا يعني** بأن كل عنصر في المجموعة لديه الفرصة ليتم اختياره.
مثلا اذا كانت عينة البحث هي Type 2 diabetes patients و كانت عدد الناس الخاضعين للتجربة 100000 مريض و نريد أن نختار عدد محدد من المئة الف فان كل واحد منها لديه نفس الاحتمال ليتم اختياره أو لا يتم . و تعد هذه العملية عملية عشوائية.

Probability sampling is the perfect way to make sampling because we give every one equal chance to be selected.

→ Allow use of inferential statistics.

A-Simple random sampling

B-Systematic sampling

C-Stratified random sampling (SRS)

D-Cluster sampling (single sage , multi stage)

A-Simple random :

Example: A list of all currently enrolled medical students at Mutah University is obtained and a table of random numbers is used to select a sample of students.

Example: A researcher obtains a list of all residential addresses in the city and uses a computer to generated a random list of homes to be included in a survey.
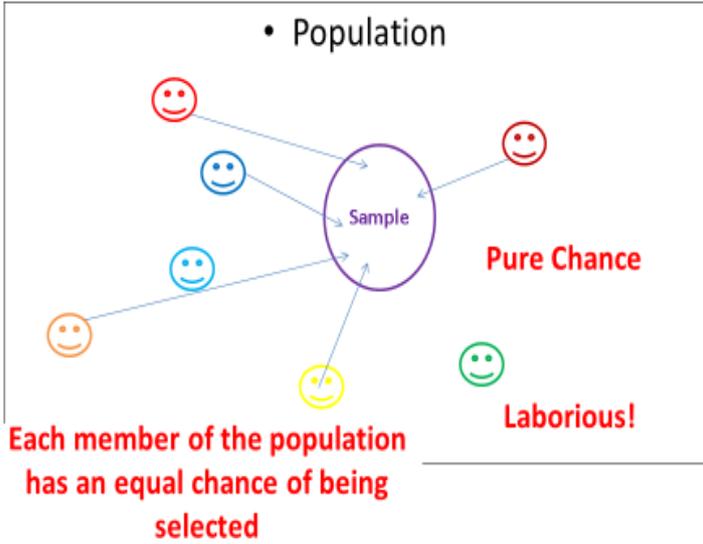
Random sampling is the **purest** form of probability sampling.
- Each member of the population has <u>**an equal and known chance**</u> of being selected.
- We use random sampling When there are <u>**very large populations**</u>, it is often 'difficult' to identify every member of the population, so the pool of available subjects becomes biased (not inclusive enough and needs additional sampling techniques).
- * **significance** is taking an Adequate number of people or items to have a right result can be inference on the population.
- You can use software, such as *minitab to generate random numbers or to draw directly from the columns

*<u>**minitab**</u> is a computer program used for **statistics**

Simple random

- Population

Sample

Pure Chance

Laborious!

Each member of the population has an equal chance of being selected

Random number table

B-Systematic :

Systematic sampling is often used instead of random sampling.

* It is also called an **Nth** name selection technique.

After the required sample size has been calculated, every **Nth** record is selected from a list of population members.

Its **advantage** over the random sampling technique is **simplicity** (and possibly cost effectiveness).

The main **Disadvantage** is the bias →As long as the list does not contain any **hidden order (bais)**, this sampling method is as good as the random sampling method.

في هذا النوع يتم اختيار نقطة البداية عشوائيا ثم نكمل بنمط ثابت (بفترات ثابتة)

اذ انه في الشكل يكون الفرق بين كل عينة اثنين أي اننا نختار كل ثالث شخص.

Systematic Random

Population

Bias! Careful!

Sample

Each member is either assembled or listed, a random start followed by selection at **equal intervals (fixed intervals)**
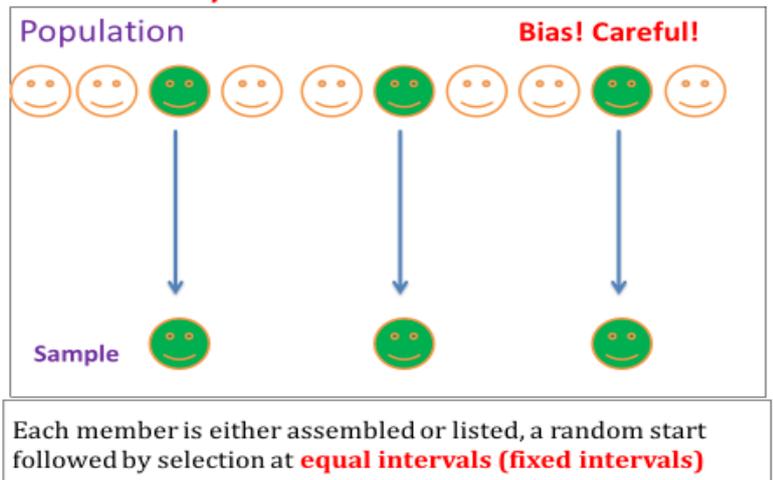
Example : We want to carry out a survey of patients with **Asthma(الربو)** attending clinic in Alkarak city. There may be too many to interview everyone, so we select a **systematic sampling**!

If 3000 patients attend the clinic in total, and we **only require** a sample of **200**, we need to:

1- Calculate the **interval** (3000/200), sampling fraction of **15**.

2- Select a **random number** between **1-15 using random tables.**

3- Suppose that number is **13** for example, we select the patient number 13 and then go on select every 15th person.

4- This should give us a total of 200 patients.

Note: A researcher selects a starting integer to base the system on.

 This number needs to be **smaller** than the population as a whole,

 e.g., he doesn't pick every 500th patient to sample for a 100 patients.

C-Stratified random sampling (SRS) :

Stratified random sampling divides the population into **smaller groups,** or **strata**, based on **shared characteristics**.

* A random sample is taken from each stratum in direct proportion to the size of the stratum compared to the population.

Proportionate and **disproportionate** stratified sampling.

The main **advantage** is presenting all of the population because we take a sample from each strata in the population.

What about statistical ** power (significance)? → the groups is not equal in the members number. so, it is not accurate

Stratified Random Sampling
Population
Group One — SRS
Group Two — SRS
Representative of the population frame!
Sample
What about statistical power??!
Group Three — SRS
Group Four — SRS

Each member is assigned to a group or a stratum, then simple random sample is selected from each stratum.
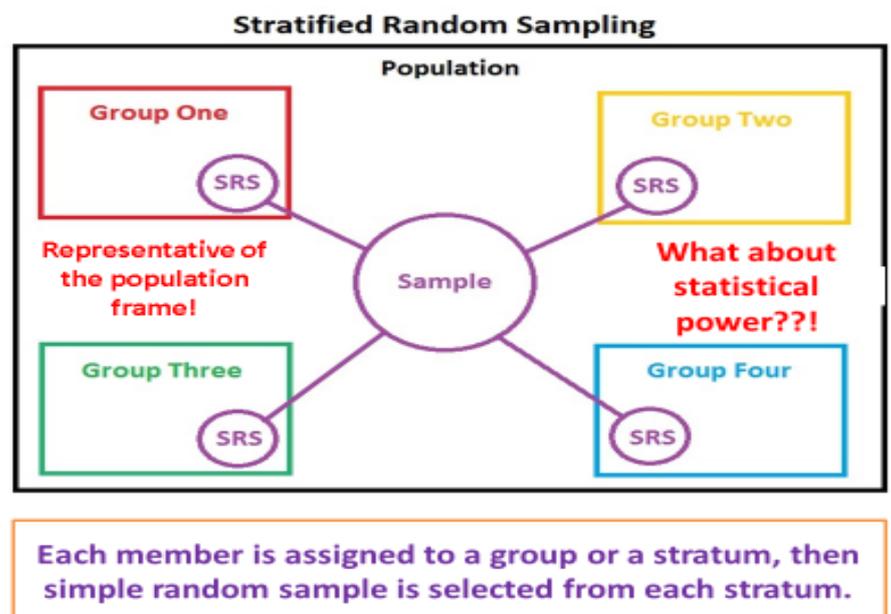
Example: A study designed to evaluate the learning process views of medical students at a major university.

The researchers want to ensure the random sample best approximates the student population including gender, undergraduate and graduate.

The total population in the study is 1,000 students and from there, subgroups are created as shown below.

Total population = 1,000

Subgroups:

Male undergraduates = 450 students (out of 1000) *or 45% of the population*

Female undergraduates = 200 students or 20%

Male graduates = 200 students or 20%

Female graduates = 150 students or 15%

Random sampling is done for each subpopulation based on its representation within the population as a whole. Since male undergraduates are 45% of the population, 45 male undergraduates are randomly chosen out of that subgroup. Because male graduates make up only 20% of the population, 20 are selected for the sample and so on. (**Proportionate**)

disproportionate there are a little number of female graduates so the result will not be significant

so we will make all number = 25 from all four groups.
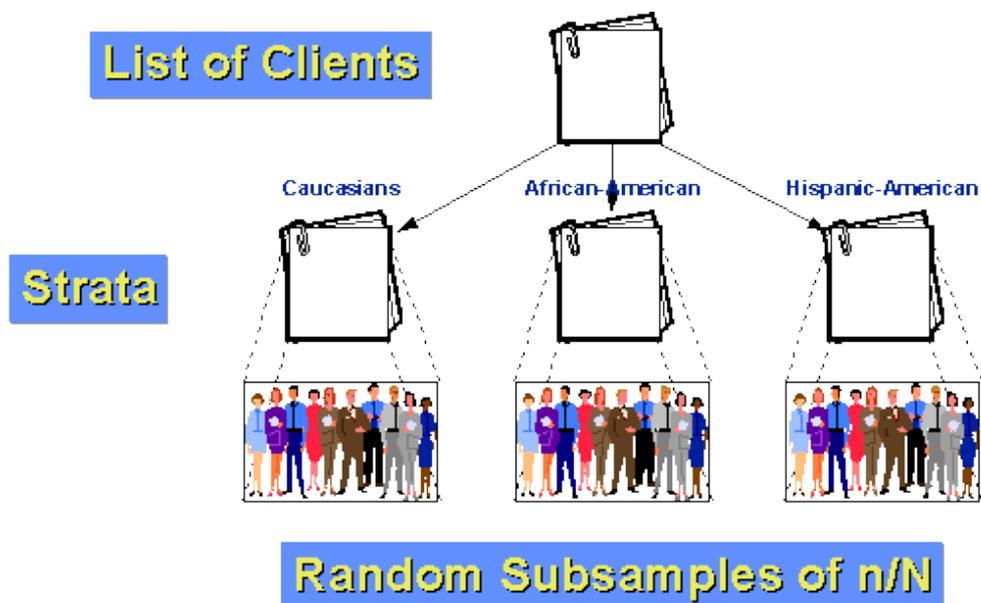
D-Cluster sampling :

We divide the population to groups

Then we chose some of these groups

Then we take all of the members

In the chosen groups

Cluster Sampling

Population

Group One

Group Two

National surveys!

Sample

Group Three

Group Four

Each member is assigned to a group or a cluster, then clusters are selected randomly where everyone in the cluster is included.

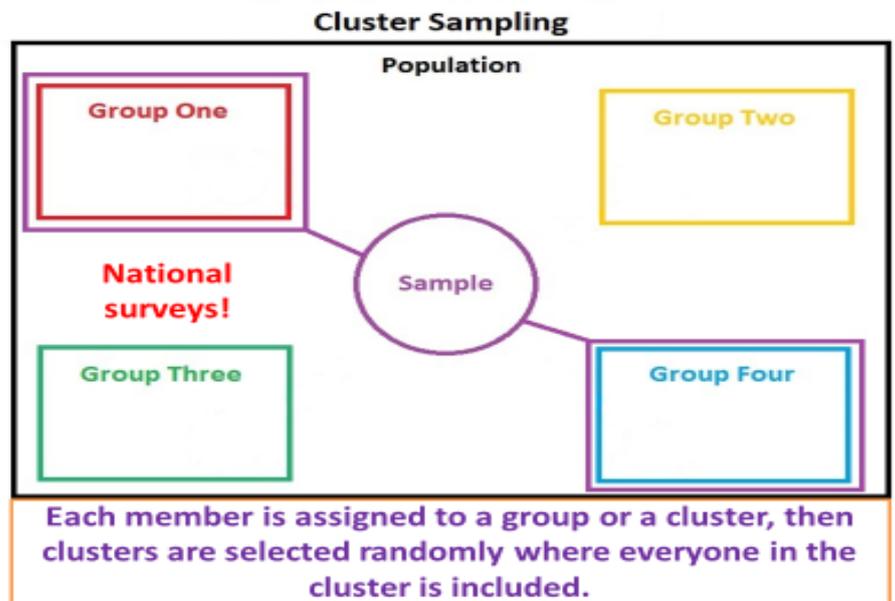- The entire population is divided into clusters or sections and then the clusters are randomly selected. All the elements of the cluster are used for sampling. Clusters are identified using details such as age, sex, location etc.

- Cluster sampling can be done in following ways:

- Single Stage Cluster Sampling
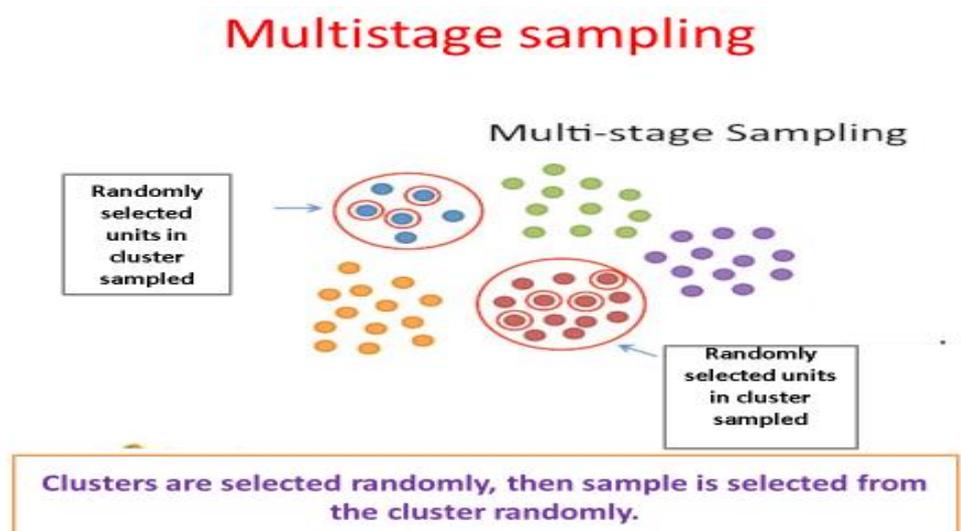
- Multi Stage Cluster Sampling

Example:

Geographical clusters are the most common.

In a city, the list of all the individual persons undergoing certain surgery in Jordan may be difficult to obtain or even may be not available but a list of all the hospitals doing that surgery in Jordan be available. So every hospital of these will be a cluster.

-Multistage sampling

First we divide the population into clusters then we choose clusters randomly and we choose

a random unit from the clusters which is chosen (could be simple random sampling, Systematic sampling or Stratified random sampling)

Multistage sampling

Multi-stage Sampling

Randomly selected units in cluster sampled

Randomly selected units in cluster sampled

Clusters are selected randomly, then sample is selected from the cluster randomly.

<h1 style="text-align:center;">Group 2 : B- non-probability</h1>

- **It does not rely on randomization. This technique is more reliant on the researcher's ability to select elements for a sample.**
- **Outcome of sampling might be biased and makes difficult for all the elements of population to be part of the sample equally.**
- **This type of sampling is also known as non-random sampling.**

Nonprobability sampling is any sampling method where some elements of the population have *no* chance of selection (these are sometimes referred to as 'out of coverage'/'undercovered'), or where the probability of selection can't be accurately determined. (non-random)

يتم اختيار العينة بناء على خصائص معينة و تعد باقي العناصر خارج نطاق التغطية أي انه لم يتم الكشف عنها

كأن نقوم باجراء بحث عن مرض معين في الاردن دون الذهاب الى جميع المحافظات فيتم الذهاب الى مناطق محدده مثل الكرك و معان لخصائص معينه و هي قربهما الجغرافي فيكون للمريض في الكرك او معان فرصة ليتم اختياره بينما لا يمتلك المريض في اربد فرصة ليتم اختياره.

Types of non-probability :

A-Quota sampling

B-Convenience or opportunistic sampling

C-Snowball or networking sampling

D-Theoretical sampling

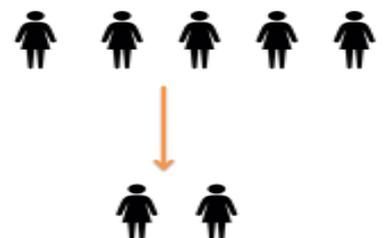
A-Quota sampling:

> **Selected criteria in advance!**
> **to increase the chance for a specific stratum to be presented**

Selected criteria in advance!

- This type of sampling depends of some pre-set standard. It selects the representative sample from the population.

- Proportion of characteristics/ trait in sample should be same as population.

- Elements are selected until exact proportions of certain types of data is obtained or sufficient data in different categories is collected.

B-Convenience or opportunistic sampling :

Convenience sampling is a non-probability sampling technique where subjects are selected because of their convenient accessibility and proximity to the researcher.

convenience sample is either a collection of subjects that are accessible or a self selection of individuals willing to participate

This method is used when the availability of sample is rare and also costly. So based on the convenience samples are selected.

Easily accessible
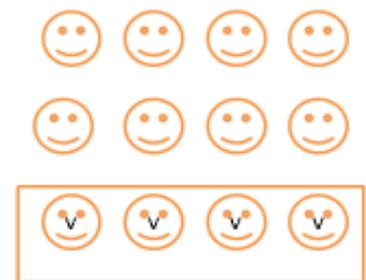

C-Snowball or networking sampling

Snowball sampling is a special nonprobability method used when the desired **sample characteristic is rare**.

It may be extremely difficult or cost prohibitive to locate respondents in these situations.

This technique relies on referrals from initial subjects to generate additional subjects.

It lowers search costs; however, it introduces bias because the technique itself reduces the likelihood that the sample will represent a good cross section from the population.
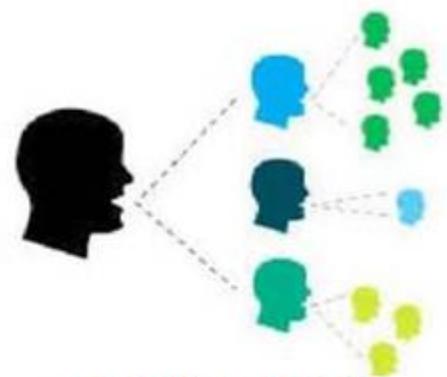
Rare or have socially undesirable characteristics!

D-Theoretical sampling(Purposive sampling):

1- Beginning of the study → decide the sample size
2- Initial theory→ is the question of the research which you want to know it.
3- Further sampling→to ensure from the results
4- Theory development→ develop the theory to have the final result from the research.

Beginning of the study   Initial theory   Further sampling   Theory development
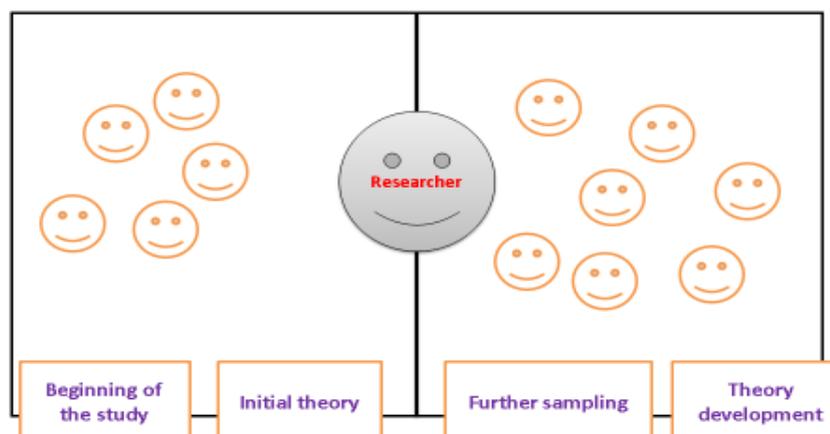
Bias in sampling

- There are five important potential sources of bias that should be considered when selecting a sample, irrespective of the method used. Sampling bias may be introduced when:

1. Any pre-agreed sampling rules are deviated from

2. People in hard-to-reach groups are omitted

3. Selected individuals are replaced with others, for example if they are difficult to contact

4. There are low response rates

5. An out-of-date list is used as the sample frame (for example, if it excludes people who have recently moved to an area

Key points to remember when deciding on sample selection are:

Always try to use a random method where possible and remember that random **doesn't mean haphazard**!

Random selection means that everyone in your sampling frame has an **equal** opportunity of being included in your study!

If you need to be able to generalize about small or minority groups and to compare those to large groups, consider using disproportionate stratified sampling , but remember to re-weight the results afterwards!