Sampling and data collection

1- Population

In statistics the term *population* is broaden used not only means the people or human residents who live in a certain region, but also refers to other group of living such as animals, insects, and plants. To embark more on this concept, the population can be used for subset specific group for instance, doctor, nurse, patients, children, elderly, and homeless people.

However, the population mean is donated by μ (mu) given by probability distribution and the standard deviation is denoted by ς (low case sigma).
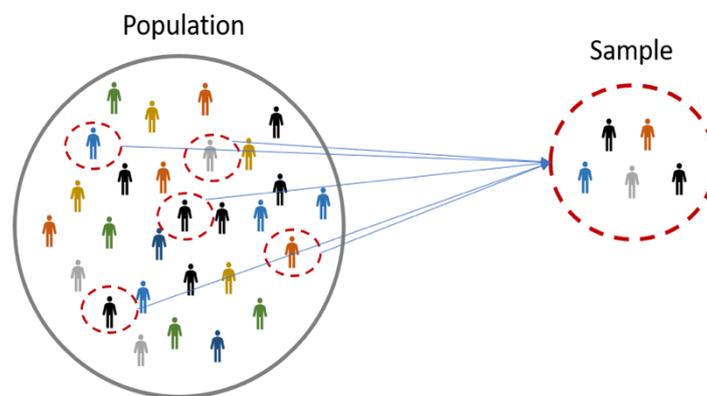

Illustration (1): Sample and Population.

2- Sample and Sampling

- Sample: is the unite or individual that may be chosen for participation in a study (people, families, household, institutions. People are referred to as subjects or participants.
- Sampling: the process of selecting a group of people, events, behaviours, or other elements can be involved for the study.
- In samples design process the questions it ask are what and who is the focus of the study, where, and when it is being done, and why.

-To clarify this by an example: focus on the cigarette-smoking behaviour (what) of the high school seniors (who) in a large metropolitan high school (where). So, this is very important to address our question in the study intend to conduct, as an essential part of the study design or plan it.

1. Types of Sampling Methods

1.1. Probability Sampling

Probability sampling means every individual in the population has a known, non-zero chance of being selected. This type of sampling is generally preferred because it provides a more accurate representation of the population and minimizes selection bias.

- **Simple Random Sampling (SRS):**
  - **Definition:** Every patient in the hospital has an equal chance of being selected. This method can be performed using random number generators or drawing names from a list.
  - **Advantages:** It's straightforward and avoids selection bias.
  - **Limitations:** Can be impractical in a hospital setting if the number of patients is very large, or the hospital doesn't have a complete list of all patients at the time of selection.
- **Systematic Sampling:**
  - **Definition:** A fixed interval is chosen, and patients are selected at regular intervals (e.g., every 10th patient).
  - **Advantages:** Easier than random sampling when dealing with large populations.
  - **Limitations:** If the list is ordered in a systematic way (e.g., by admission date), it may introduce bias if there's a pattern in the order that correlates with the condition being studied.
- **Stratified Sampling:**
  - **Definition:** The population is divided into subgroups (strata) based on specific characteristics (e.g., age, sex, disease type), and then a random sample is taken from each subgroup.
  - **Advantages:** It ensures that all relevant subgroups are represented and provides more precise estimates.
  - **Limitations:** Requires knowledge of the strata characteristics beforehand and can be more complex to manage.
- **Cluster Sampling:**
  - **Definition:** The hospital is divided into clusters (e.g., by department, ward, or floor), and then a random sample of clusters is selected. Afterward, all patients within the selected clusters are sampled.
  - **Advantages:** Useful when it's difficult or expensive to obtain a complete list of patients.
  - **Limitations:** Less precise than simple random sampling because within-cluster variability may be lower than the variability in the general population.

1.2. Non-Probability Sampling

Non-probability sampling does not give every patient in the population an equal chance of being selected, which can introduce bias. However, it may still be appropriate in certain circumstances.

- **Convenience Sampling:**
  - **Definition:** Patients are selected based on their availability or accessibility (e.g., patients who are present in the hospital at the time of the study).
  - **Advantages:** Easy to implement and cost-effective.

- **Limitations:** Prone to selection bias, as patients who are available may not represent the entire population (e.g., patients with certain conditions may be overrepresented).
- **Judgment or Purposive Sampling:**
  - **Definition:** The researcher selects specific patients based on certain characteristics (e.g., patients who have a specific diagnosis, age group, or outcome).
  - **Advantages:** Focused on specific outcomes or characteristics that the researcher is interested in studying.
  - **Limitations:** Can introduce bias and limit the generalizability of findings because the sample may not represent the larger population.
- **Quota Sampling:**
  - **Definition:** The researcher ensures that the sample includes a certain number of patients from each subgroup (e.g., 50 men, 50 women). The selection within each subgroup may be done non-randomly.
  - **Advantages:** Ensures that certain subgroups are represented in the sample.
  - **Limitations:** Non-random selection within subgroups can introduce bias.

2. Ethical Considerations in Patient Sampling

In a hospital setting, ethical considerations are paramount when selecting patients for a study. These include:

- **Informed Consent:** Patients must be informed about the study, including its purpose, methods, and any potential risks. Their consent must be freely given, and they should understand that they can withdraw at any time without affecting their care.
- **Confidentiality:** Patients' personal information and medical history must be kept confidential. Only anonymized or coded data should be used for research purposes.
- **Vulnerable Populations:** Special considerations should be taken when dealing with vulnerable populations, such as children, the elderly, or those with cognitive impairments. In these cases, additional safeguards may be needed, such as obtaining consent from a legal guardian.
- **Risk vs. Benefit:** The potential benefits of the research must outweigh any risks posed to the patients. The study should aim to improve patient care or contribute to scientific knowledge that could benefit the population at large.

3. Sample Size Determination

Determining the appropriate sample size is crucial to ensure that the study has sufficient statistical power to detect meaningful differences or associations. This can be calculated based on:

- **Effect size:** The magnitude of the difference you expect to find (e.g., treatment effect, disease prevalence).
- **Significance level (alpha):** The probability of making a Type I error (usually set at 0.05).
- **Power (1 - beta):** The probability of detecting a true effect (usually set at 80%).
- **Variance:** The expected variability in the data.

Note: Sample size calculations can be done using statistical software or formulas designed for specific study designs.

4. Data Collection Methods

Once the sample is selected, various data collection methods may be used, depending on the study objectives. These include:

- **Medical Records Review:** Data can be obtained by reviewing patient charts for information about diagnoses, treatments, outcomes, and demographics.
- **Surveys and Questionnaires:** Structured interviews or surveys may be used to gather information directly from patients, especially for studies that involve subjective outcomes such as quality of life or patient satisfaction.
- **Clinical Observations:** Researchers or clinicians may make direct observations about a patient's clinical status, behaviors, or physical symptoms.
- **Laboratory Tests:** In studies examining biomarkers or other biological factors, blood, urine, or tissue samples may be collected for analysis.