Research Data collection & Sampling

Outlines:

- **Data Collection**
- **Process of Data Collection**
- **Methods of Data Collection**
- **Measurement**
- **Sampling**

Definition of Data Collection:-

Data collection:

Is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions and evaluate outcomes. Data collection is the systematic approach to gathering and measuring information from a variety of sources to get a complete and accurate picture of an area of interest.

Process of Data collection:-

1)Clarify your data collection goals.

- What is your data collection goal ? the main consideration of data collection process is answering the research question or research hypothesis.

2)Develop operational definitions and procedures.

- Here we need to be very clear as to what we are measuring, how it is to be measured, and who is to measure it. Often times we will employ sampling in which case we need to define a sampling plan

3)Validate the measurement system.

- In order for research data to be of value and of use, they must be both reliable and valid.
- **Reliability** refers to the repeatability of finding. If the study were to be done a second time, would it yield the same results? If so the data are reliable. If more than one person is observing behavior or some event, all observers should agree on what is being recorded in order to claim that the data are reliable.
- **Validity** refers to the credibility or believability of the research which the instruments or procedures used in the research what they were supposed to measure.

4)Begin data collection.

- Using all the knowledge from the previous steps we now go off and collect our data.

5)Continue improving measurement system and ensure people are following the data collection guidelines

Methods of Data Collection:-

1) Observation:-

Defined as is way of gathering data by watching behavior, events or noting physical characteristics in their natural setting

- **Types of observation:-**

1)**Overt**: Everyone knows they are being observed.

2)Covert: No one knows they are being observed and the observer is concealed.

3)Direct: When you watch interactions, processes, or behaviors as they occur

4)Indirect: Indirect observations are when you watch the results of interactions, processes, or behaviors.

- **When should you use observation for collecting data?**

1-When you are trying to understand an ongoing process or situation. Through observation you can monitor or watch a process or situation that you are evaluating as it occurs.

2-When you are gathering data on individual behaviors or interactions between people. Observation allows you to watch peoples' behaviors and interactions directly or watch for the results of behaviors or interactions

3-When you need to know about a physical setting. Seeing the place or environment where you're something takes place can help increase understanding of the event, activity, or situation you are evaluating. **For example,** you can observe whether a classroom or training facility is conducive to learning.

4-When data collection from individuals is not a realistic option: If respondents are unwilling or unable to provide data through questionnaires or interviews, observation is a method that requires little from the individuals for whom you need data

Advantages of observation	Disadvantages of observation
Access to situation and people where questionnaire are impossible to use	**Can be view as too subjective**
Access to people in real life	**Time consuming**
Good for explaining for meaning and context	**Depend on the role of researcher**

Can be strong on validity and in-depth understanding	**Overt: may affect situation and validity of finding**
	Covert: ethical principles contravened
	High potential for role conflict for practitioner researcher

2) Interviewing

The collection of data by asking people questions and following up or probing their answers.

- ***Telephone interview***

Advantages:

1)Less time consuming
2)Less expensive
3) Researcher has ready access to anyone who has landline telephone

Disadvantages:

1)As face to face interview the response rate is not a high.
2)The sample may be biased as only those people who have landline phone are contacted

- ***Face –to-Face interview:-***

Advantages:

1)Enable the researcher to establish the rapport with potential participant and therefore gain their cooperation
2)Yield the highest response rate in survey research
3)Allow the researcher to clarify the ambiguous answer and when appropriate**.**

Disadvantages:

1)Impractical when large sample are involved.
2) Can be time consuming and expensive

- ***Computer Assist Personal Interviewing (CAPI):***

Form of personal interviewing but instead of completing questionnaire, the interviewer bring along laptop or hand held computer to inter information directly into database

Advantages:

1)Save time involved in processing the data.
2) Save the interviewer from caring around hundreds of questionnaire

Disadvantages:

1)Can be expensive to set up.
2) Require the interviewer have computer and typing skills

3)Questionnaire:-

It Is a data collection tool in which written questions are presented that are to be answered by the respondents in written form

Questionnaire can take several forms as:-

- ***Closed-ended questionnaire.***

For example:-

1. Gender Status?(a) Male (b) Female

2. Marital Status?(a) Married (b) un-married

- ***Open-ended questionnaire.***

For Example:-

- **What do you know about poverty?**
- **What are the types of poverty?**

Types of questionnaire

1) Hand Delivered Questionnaire:-

- This is a type of questionnaire in which the investigator himself go to the field and hand over the pre written questions to the respondents.
- They only tick mark, the correct answers in front of the investigator. It is also called direct questionnaire because the researcher directly distributes the questionnaire among the respondents.

Advantages

- **The researchers have close contact with respondents.**
- **Difficult questions are explained by the researcher to the respondents.**
- **He explains the purpose of the study**

Disadvantages

- **It is more expensive and costly.**
- **It more time consuming**

2) Mailed Questionnaire:-

- In that type the respondents are living in for-flung areas at a distance and the questionnaire is sent to them by post, they fill it and return back to the researcher or concerned department. A particular guide line or instructions list is attached to the questionnaire for the respondent's guidance

Advantages

- **It is commonly used.**
- **It is useful for the researcher.**
- **It is very easy and simple.**
- **It saves time and money**

Disadvantages

- **Lack of returns.**

- **Research take time due to careless and laziness of the respondents.**
- **Lack of skilled respondents.**
- **Errors may occur due to misunderstanding of respondents**

4) Focus group:-

- A focus group is a market research method that brings together 6-10 people in a room to provide feedback regarding a product, service, concept, or marketing campaign. A trained moderator leads a 30-90-minute discussion within the group that is designed to gather helpful information. The moderator arrives with a set list of 10-12 questions that will be shared with the group during their time together that are designed to elicit thoughtful responses from all the participants. The moderator's goal is to hear from everyone and to encourage many different opinions and ideas to be shared

Advantages:-

- **It is an inexpensive and fast method of acquiring valuable Data.**
- **Informants can build on the answers of others**
- **The researcher can clarify clashes among participants and ask about these diverse opinions.**
- **All individuals along with the researcher have a chance to ask questions, and these will produce more information when compared with individual interviews.**

Disadvantages:-

- **The researcher has trouble controlling discussion and managing the process in comparison to individual interview.**
- **Recording data can present difficulties; it is actually not possible to record when so many participants are speaking at the same time. Also tape recorders may record just those who are closer.**
- **Data analysis could be time consuming and challenging task.**

Definition of measurement

- Is the process of assigning numbers to variables it include counting ; ranking and comparing objectives or events some qualitative studies gather data in narrative form and numbers are not associated with these data so these data not included in the concept of measurement.

Level of measurement

- **Not all data is created equally .it is helpful to classify data sets by different criteria .some is qualitative, and some is quantitative.**
- **Some data sets are continuous and some are discrete.**

- **Another way to separate data is to classify it in to four levels of measurement: nominal, ordinal, interval and ratio. we will look at each of these levels of measurement.**

Nominal Level OF Measurement

- The nominal level of measurement is the most primitive method of classifying information, Nominal means "in name only" and that should help to remember what this level is all about .Nominal data deals with (names, labels or categories. Data at the nominal level is qualitative .Colors of eyes and yes or no responses to a survey, all deal with nominal level. It characterized by variable s that are discrete and noncontiguous.
- These variables categorical and include such examples as sex (male and female), marital status (married, unmarried) blood types (o, B, A, AB) and health status (sick, well)

Ordinal Level of Measurement

- The next level is called the ordinal level of measurement .it contains all of information captured in the nominal scale but it also ranks data from lowest to highest .Rather than simply categorize data by placing an object either into or not into a category .ordinal data give you some idea where data lie in relation to each other .Data at this level can be ordered but no differences between the data can be taken that are meaningful.
- Here you should think of things like a list of the ten cities to live. The data , here ten cities ,are ranked from one to ten ,but differences between the cities don't make much sense .There is no way from looking at just the rankings to know how much better life is in city number 1 than city number 2. As with the nominal level, data at the ordinal level should not use in calculations.
- **Interval Level Of Measurement**

- Unlike the nominal that simply places objects into or out of a category or the ordinal level that rank orders objects, the interval level indicates the distance one object from another. interval level deals with data that can be ordered, and in which differences between the data does make sense .

- Data at this level does not have a starting point .interval level refers to the complexity of the statistical techniques that can be used to analyze data. Variables with in this level of measurement are assessed incrementally, and the increments are equal .Many statistical technique can be used to analyze interval variables

- The Fahrenheit and Celsius scales of temperatures are both examples of data at the interval level of measurement .You can talk about 30 degrees being 60 degrees less than 90 degrees, so differences can make sense .HOWEVER **0 DEGREES** (In both scales) cold as it may be does not represent the total absence of temperature .Data at the interval level can be used in calculations. HOWEVER ,data at this level does lack one type of comparison .Even though 3x30=90,it does not correct to say that 90 degree Celsius is three times hot as 30 degree .

- **RATIO Level of Measurement**

- The level that contains the richest information about an object is ratio level. The ratio level contains all of information of the previous three levels plus it contains an absolute zero point .Due to presence of a zero, it now makes sense to compare the ratios of measurement .Phrases such as "four times " twice " are meaningful at the ratio level .Distances .in any system of measurement give us data at the ratio level .

- A measurement such as 0 feet does make sense, it represent no length. furthermore 2 feet is twice as long as one feet .so ratios can be formed between data .At the ratio level of measurement , not only can sums and differences be calculated ,but also ratios .One measurement can be divided by any nonzero measurement ,and a meaningful number will result .

- **Sampling:**

- In clinical research, we define the population as a group of people who share a common character or a condition, usually the disease.

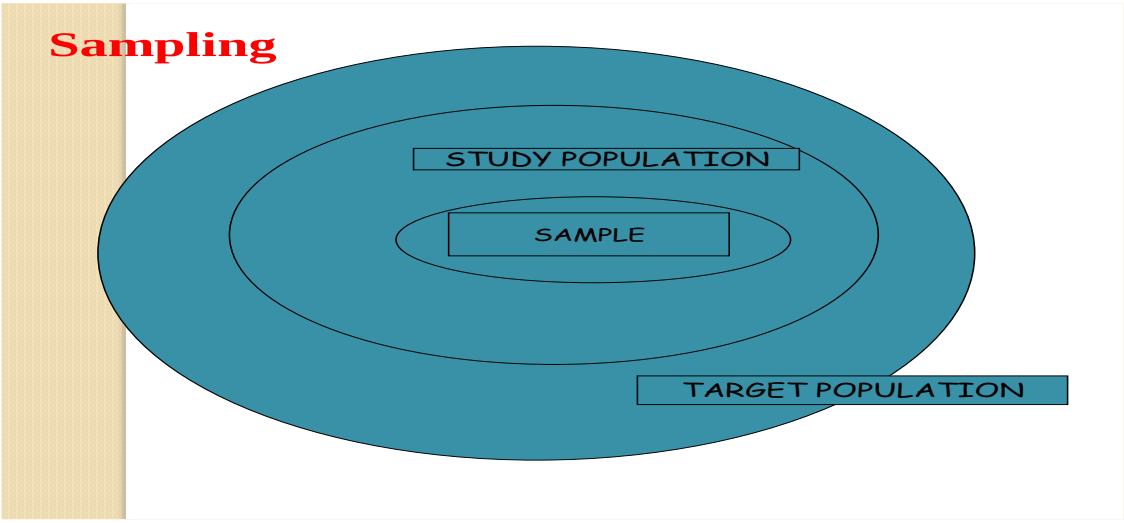
- If we are conducting a study on patients with ischemic stroke, it will be difficult to include the whole population of ischemic stroke all over the world.

- It is difficult to locate the whole population everywhere and to have access to all the population.

- The practical approach in clinical research is to include a part of this population, called “sample population”. The whole population

is sometimes called “target population” while the sample population is called “study population.

- When doing a research study, we should consider the sample to be representative to the target population, as much as possible, with the least possible error and without substitution or incompleteness. The process of selecting a sample population from the target population is called the “sampling method”.



- **Sampling**

- -A major purpose of doing research is to infer or generalize research objectives from a sample to a larger population.
- - The process of inference is accomplished by using statistical methods based on probability theory.
- Definition of Sample:
- A sample is a subset of the population selected, which is an unbiased representative of the larger population.
- Studies that use samples are less-expensive, and study of the entire population is sometimes impossible.
- Thus, the goal of sampling is to ensure that the sample group is a true representative of the population without errors.

Methods of sampling

1- probability (random) sampling

2- Non probability (Non random) sampling

- To ensure reliable and valid inferences from a sample, probability sampling technique is used to obtain unbiased results.

- In probability sampling methods all subjects in the target population have equal chances to be selected in the sample.

- In non-probability sampling methods
- the sample population is selected in a non-systematic process that does not guarantee equal chances for each subject.
- in the target population .Samples which were selected using probability sampling methods are more representatives of the target population.

The four most commonly used probability sampling methods are:

- Simple random sampling,
- Stratified sampling
- Systematic sampling,
- Cluster sampling

1- Probability Sampling:

- ***1-1-Simple random sampling***

This method is used when the whole population is accessible and the investigators have a list of all subjects in this target population.

The list of all subjects in this population is called the "sampling frame". From this list, we draw a random sample using lottery method or using a computer generated random list.

1.2. Stratified random sampling

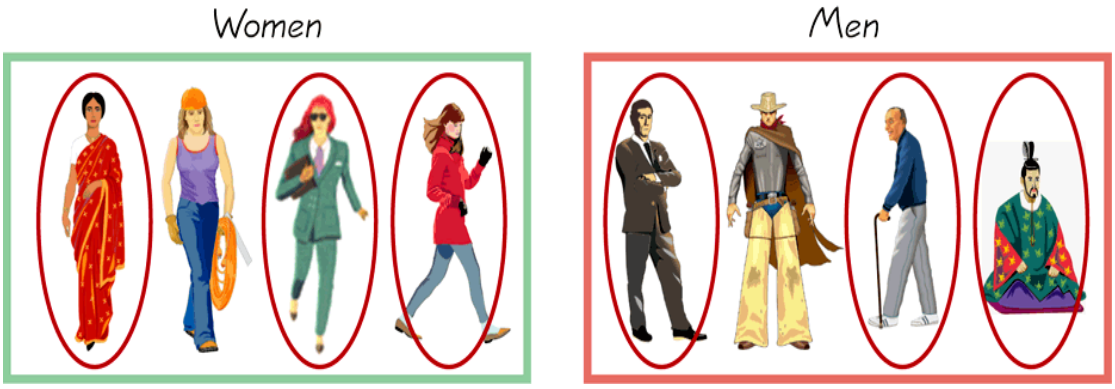
This method is a modification of the simple random sampling therefore, it requires the condition of sampling frame being available, as well.

in this method, the whole population is divided into homogeneous strata or subgroups according a demographic factor (e.g. gender, age, religion, socio-economic level, education, or diagnosis etc.). Then, the researchers select draw a random sample from the different strata .

The advantages of this method are:

(1) It allows researchers to obtain an effect size from each strata separately, as if it was a different study.

draw a random sample from the different strata



Therefore, the between group differences become apparent.

(2) It allows obtaining samples from minority/ under-represented populations. If the researchers used the simple random sampling, the minority population will remain underrepresented in the sample, as well. Simply, because the simple random method usually represents the whole target population. In such case, investigators can better use the stratified random sample to obtain adequate samples from all strata in the population.

1-3 :Systematic random sampling (Interval sampling)

In this method, the investigators select subjects to be included in the sample based on a systematic rule, using a fixed interval.

For example: If the rule is to include the last patient from every 5 patients. We will include patients with these numbers (5, 10, 15, 20, 25, ...etc.). In some situations, it is not necessary to have the sampling frame if there is a specific hospital or center which the patients are visiting regularly.

In this case, the researcher can start randomly and then systemically chooses next patients using a fixed interval.

1-4:Cluster sampling (Multistage sampling)

It is used when creating a sampling frame is nearly impossible due to the large size of the population. In this method, the population is divided by geographic location into clusters. A

list of all clusters is made and investigators draw a random number of clusters to be included. Then, they list all individuals within these clusters, and run another turn of random selection to get a final random sample exactly as simple random

sampling. This method is called multistage because the selection passed with two stages:

- ✓ firstly, the selection of eligible clusters, then, the selection of sample from individuals of these clusters.

An example for this, if we are conducting a research project on primary school students from Egypt . It will be very difficult to get a list of all primary school students all over the country. In this case, a list of primary schools is made and the researcher randomly picks up a number of schools, then pick a random sample from the eligible schools.

2-Non-probability sampling method

2.1. Convenience sampling

Although it is a non-probability sampling method, it is the most applicable and widely used method in clinical research.

-In this method, the investigators enroll subjects according to their availability and accessibility. Therefore, this method is quick, inexpensive, and convenient.

- It is called convenient sampling as the researcher selects the sample elements according to their convenient accessibility and proximity.

For example: assume that we will perform a cohort study on Egyptian patients with Hepatitis C (HCV) virus.

The convenience sample here will be confined to the accessible population for the research team. Accessible population are HCV patients attending in Zagazig University Hospital and Cairo University Hospitals.

Therefore, within the study period, all patients attending these two hospitals and meet the eligibility criteria will be included in this study.

2-2 :Judgmental (Purposive) sampling

In this method, the subjects are selected by the choice of the investigators. The researcher assumes specific characteristics for the sample (e.g. male/female ratio = 2/1) and therefore, they judge the sample to be suitable for representing the population. This method is widely criticized due to the likelihood of bias by investigator judgment.

3-2:Snow-ball sampling

This method is used when the population cannot be located in a specific place and therefore, it is different to access this population. In this method, the investigator asks each subject to give him access to his colleagues from the same population. This situation is common in social science research,

- for example, if we running a survey on street children, there will be no list with the homeless children and it will be difficult to locate this population in one place e.g. a school/hospital. Here, the investigators will deliver the survey to one child then, ask him to take them to his colleagues or deliver the surveys to them.

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:

- Any pre-agreed sampling rules are deviated from
- People in hard-to-reach groups are omitted
- Selected individuals are replaced with others, for example if they are difficult to contact
- There are low response rates
- An out-of-date list is used as the sample frame (for example, if it excludes people who have recently moved to an area)