Biostatistics

Doctor 2017 | Medicine | JU

Number >>	2
Doctor	Dr. Hamza Alduraidi
Done By	Hind Bilal
Corrected By	Rawan Almujaibel

In the previous lecture, we said that we conduct scientific research in order to understand the phenomena around us for better understanding of the universe.

When we do scientific research, first thing we do, we collect **Data** which is the simplest and smallest shape of information (building blocks for understanding) represented in numbers, characters or words from people. We can obtain the data from people by **analysis of records** from the past, **surveys** that we did by ourselves**, counting** people with a certain disease for certain event and from **experiments** which are the most respective source of data to run our own experiment. Also from **reports** (mean studies have been done by other people in the past). Regardless of the source of the data we then use it to draw our conclusion and to find our results.

From the slide:

- Data are numbers which can be measurements or can be obtained by counting
- Biostatistics is concerned with the interpretation of the data and the communication of information about the data.

Actually, the process of understanding the universe around us refers to as **Logical reasoning.**

From philosophers of the Greeks until today, the scientists are divided into two groups regarding to their type of logical reasoning and understanding of the universe. As a result of this we get 2 types of research which we discussed in the previous lecture.

They are 2 methods of research, which are **quantitative** and **qualitative researches**.

In this lecture we will be talking about the origin of these methods and their philosophy briefly.

NOTE: Terminological difference: in quantitative research the sample is called (subject), whereas in qualitative it's called (participant).

Logical Reasoning: Is a method of knowing combines experience, intellectual faculties, and formal systems of thought.

Scientists and Philosophers throughout the history are divided into having diffirent methods of logical reasoning:

Inductive Reasoning: Is the process of developing generations from specific observations. E.g., a nurse may observe the anxious behavior of (specific) hospitalized

children and conclude that (in general) children's separation from their parents is stressful.

Or people start from specific observations and individuals then they build a conclusion of the bigger picture (generalization)

Deductive Reasoning: Is the process of developing specific perditions from general principles.

Which means they start from putting general rules for all people and then they can access to the individuals they are interested in.

Both of them are equally correct and important, but each has a specific way of understanding. NO ONE ALONE IS ENOUGH; BOTH ARE A COMPLEMENTARY to each other.

Slides:

Inductive: from particular to generalization

Deductive: from general to specific, the conclusion is valid not true.

Deductive Reasoning – A type of logic in which one goes from a general statement to a specific instance.

The classic example:

Socrates said: All men are mortal. (Major premise)

I am a man. (Minor premise)

Therefore, Socrates is mortal. (Conclusion)

Inductive Reasoning, involves going from a series of specific cases to a general statement. The conclusion in an inductive argument is never guaranteed.

Example: What is the next number in the sequence 6, 13, 20, 27…?

(We examine each number to find the general pattern.

PHILOSOPHIES OF RESEARCH: All researches are based on philosophical beliefs about the world's view or paradigm (thoughts).

To reach both knowledge (science) and philosophy (wisdom), Researchers and scientists must use one of these 2 methods of research:

The first method is about a philosophy school (school of thought), it starts from depending on the philosophical believes about the worlds new pyramiding, which is the **Naturalistic approach** of philosophy ,they depend on the **Inductive reasoning** which is a **Qualitative research**, they believe that people are not the same and each individual is unique , so we must understand individuals separately (so for example I do an interview with each one and ask open-ended questions so they could express their own unique experience).

Qualitative = Naturalistic

The other type is not interested in personal differences of people, and in what is true for each personal, it is called **Positivism approach** .People are seen as an entire population and this approach make their conclusions correct depending on the whole population. They survey is large group of people.

Quantitative = positivism

Methods of Research:

Research methods are the techniques used by researchers to structure a study and to gather and analyze information relevant to the research question

n. **Quantitative research**, which is most closely allied with the **positivist** tradition.

Qualitative research, which is most often associated with **naturalistic** inquiry.

2 major paradigms (means big school of thoughts):

1. **Positivism** (also, called modernism, logical positivism)

-There is a reality out there that can be studies and known

-Phenomenon is not haphazard or random. Also, they do care about the big picture (all) rather than the individual (all of them have the same truth [T]).
-Highly objective in pursuit of knowledge.
-Rooted in 19th century thought. Guided by such philosophers as Mill, Newton, & Locke.

2. **Naturalism** (also, called phenomenological, constructive)

-Putting structures and ideas in a new ways.
-Reality is not fixed.
-Subjectivity is the primary way of understanding the phenomenon of interest. They believe that each individual has their own understanding of the universe.

-Began as a countermovement to positivism with writers such as Weber & Kant.

-KANT came up with; what is true to everybody is not necessary true to each individual (has their own truth- it come as [t]).

Positivist Paradigm	Naturalistic Paradigm
• **Deductive processes**	• **Inductive processes.**
• Emphasis on discrete, specific concepts.	• Emphasis on entirely of some phenomenon, holistic.
• Verification of researchers' hunches.	• Emerging interpretations grounded in participants' experiences.
• Fixed design.	• Flexible design.
• Tight controls over contexts.	• Context-bound.
• Emphasis on measured, **quantitative information**; statistical analysis.	• Emphasis on narrative information; **qualitative analysis**.
• **Seeks generalizations.**	• **Seeks patterns.**

s and school

them are valid and each of them has its own usage.

From here until the end of the course we are going to talk about quantitative research exclusively.

FROM the slide:

Quantitative research:

- Associated with positivism.
- Use a scientific approach.
- Mainly to understand the phenomenon of intreset.
- Use deductive reasoning to general hunches that are tested in the real world.
- Use mechanisms designed to control the study.
- Researchers gather empirical evidence.
- Must focus on human beings, who are inherently complex and diverse.

Qualitative research:

- Mainly to understand the human experience.
- Rich and in-depth information.
- Flexible.
- Associated with naturalistic paradigm.

- Use inductive reasoning.

……

Assumption: Methodologic “How is knowledge obtained?”

Population and sampling

It is impossible to the whole population to be fully studied, as for example in the previous lecture to study every single adult woman and see if she had a skin cancer after using a sunscreen. That’s why we need to recruit samples from within those populations, so what is a sample?

A sample is a group of individuals that represent a specific demographic of population and it is small enough so we can manage it.

A population is the collection or set of all the values that a variable may have. The entire category is under consideration.

A sample is a part of a population. The portion of the population that is available or it is being available for analysis.

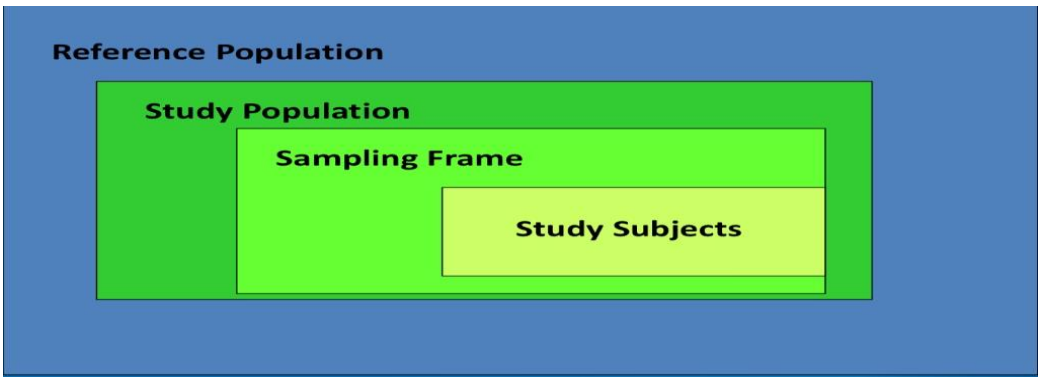
Sampling: is the process of going to a certain area and selecting a portion of the population.

Representativeness: the key characteristic of the sample is close to the population.

Sampling bias: means excluding any subject without any scientific rational. Or not based on the major inclusion and exclusion criteria. (it is also a systematic error that we must avoid)

Before sampling, you have to consider a few things. Firstly, what is the bigger population (reference) that you are trying to make conclusions about, ex: All college students in Jordan are the **reference population** and we are interested in studying about the students in the University of Jordan called **Study population**. We need to pick the target population, let’s say we want to study about medical students, we pick the Faculty of Medicine where our target could be found “sampling frame”. **Sampling frame**: where (the geographic area or place) I want to collect or found the sample .If we

are interested in second year medical students then we need to exclude other students from the sampling frame (called our sample) or it's called **accessible population**. And each selected individual in our sample called **sample element.**



Note: in the qualitative research callled participantes instead of sample element.

Questions to be considered:

- Reference population – to whom are the results going to be applied?
- What is the group of people from which we want to draw a sample (study population)?
- How many people do we need in our sample (Sample Size)?
- How will these people be selected (Sampling Method)?

From the slide:

- Element: the single member of the population (population element or population member are used interchangeably).
- Sampling frame in the listing of all element of a population, i.e., a list of all medical students at the University of Jordan, 2014-2016.

Now the process of selecting the sample members from the sampling frame is called the sampling process, there are 2 types**: Probability sampling** and **Nonprobability sampling,**

Probability sampling means that the chance for each population member to be selected is equal (randomly).

Probability Sampling Methods:

• Involves the use of random selection process to select a sample from members or elements of a population. They are:

1. Simple Random Sampling
2. Systematic sampling.
3. Stratified sampling.
4. Multistage sampling.
5. Cluster sampling

•Involves random selection procedures to ensure that each unit of the sample is chosen based on chance

• All units of the study population should have an equal or at least a known chance of being included in the sample

• Requires a sampling frame

• Listing of all study units.

The 5 types for probability sampling are:

a) **Simple random sample:** the simplest one (choosing randomly and blindly).For example, we choose 500 students from 2500 by putting all the names of people who belong to our sampling frame(medical students in JU) in a list and picking randomly the number of people you want. There are tables in certain statistical websites that give you sets of numbers that are totally random. But the problem here that we can't find a list for all the individuals' names of our population we want.

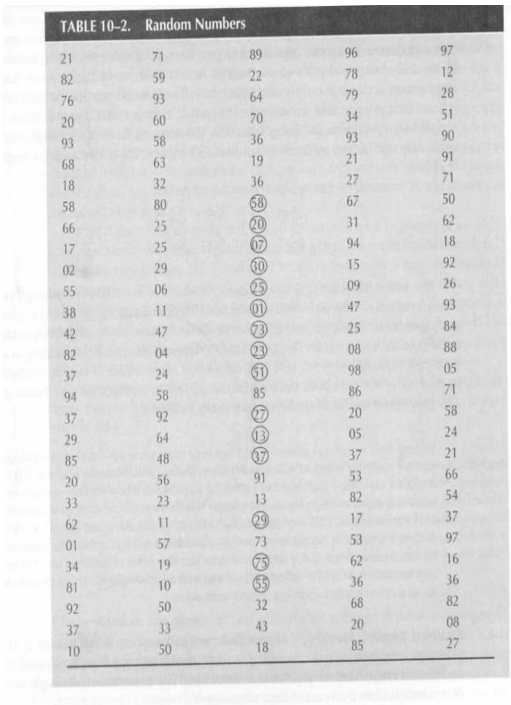


TABLE 10–2. Random Numbers

21	71	89	96	97
82	59	22	78	12
76	93	64	79	28
20	60	70	34	51
93	58	36	93	90
68	63	19	21	91
18	32	36	27	71
58	80	(58)	67	50
66	25	(20)	31	62
17	25	(07)	94	18
02	29	(30)	15	92
55	06	(25)	09	26
38	11	(01)	47	93
42	47	(73)	25	84
82	04	(23)	08	88
37	24	(51)	98	05
94	58	85	86	71
37	92	(27)	20	58
29	64	(13)	05	24
85	48	(37)	37	21
20	56	91	53	66
33	23	13	82	54
62	11	(29)	17	37
01	57	73	53	97
34	19	(75)	62	16
81	10	(55)	36	36
92	50	32	68	82
37	33	43	20	08
10	50	18	85	27

From the slide:

•Make a numbered list of all units in the population.

• Decide on the sample size.

• Select the required number of sampling units using the lottery method or a random number table.

b) **Systematic random sample:** to make things more equal and to be more conservative in giving everybody an equal chance. For ex: from the entire 2500 I need 500 (so I have to choose from every 5 persons only one), let say we will start and select person number 3 then we count 5 then we choose person number 8 and so on until we choose all 500 students that is will be our sample. So basically we choose the interval that we will go for it (such as every 5 in pervious example) and we choose randomly the starting point (number 3 as above example).

Somebody would say what about other numbers they didn't get a chance to be selected?

No, they had a chance to be selected as we choose the starting point randomly; it could be any number such as 2, 4, and 7, so on rather than number 3, so we guarantee that each individual has the equal chance to be selected.

From the slide:

• Individuals are chosen at regular intervals from the sampling frame.

• Ideally, we randomly select a number to tell us the starting point.

• Sampling fraction = Sample size/ Study population

• Interval size= study population /Sample size

c) **Stratified sampling (it comes from "stratum" which is a laten word):**

For example, our research is relevant to medical students from 1st, 2nd, 3rd year… and so on, each year has 500 students. From all these we want to choose 100 students, from the first strata (1st year) we choose a random section, and we do the same thing for the other students in different years.

From the slide:

• If we have study units with different characteristics which we want to include in the study then the sampling frame needs to be divided into strata according to these characteristics

• Ensures that proportions of individuals with certain characteristics in the sample will be the same as those in the whole study population

• Random or systematic samples of predetermined sample size will have to be obtained from each stratum based on a sampling fraction for each stratum

d) **Multistage sampling:** from all the medical students in JU (2500 students). At the first stage, we divide them into strata (1st, 2nd, 3rd and so on) then we choose the 100 students from each strata by using systematic sampling (not randomly). So by this way, we used more than one type, first the stratified then the systematic sampling and some people use 3 or even 4 ways in their study (the best way to have equal chance of being selected in a fair way).

From the slide:

- Involves more than one sampling method.
- Is therefore arried out in phases.
- Dose not require a initial sampling frame of whole population.
- NEED TO KNOW SAMPLING FRAME OF CLUSTERS E.G. PROVINCES.
- Require sampling frames of final clusters final clusters.
- Applicable to community based studies e.g. interviewving people from different villages selected from different areas, selected from different districts, provinces

e) **Cluster sampling:** (like the stratified sampling) the 2500 students that I want to choose from, they are divided into sections"clusters"(for example 2nd year students have section1, section2, section3 and so on). And each section has 100 students, so from all the years I choose 5 whole sections and get the 500 students I want.

Another example to justify it more,

I categorize students into 2 clusters (females and males) then I pick all the females who have the odd number for example.

From the slide:

- Selection of study units (clusters) instead of the selection of individuals.
- All subjects/units in the cluster who meet the criteria will be sampled.
  -Clusters often geographic units.
  -E.g. schools, villages etc.

- Usually used in interventional studies.

  -E.g. assessing immunization coverage

- **Advantages:**

-sampling frame is not required in this case

-Sampling study population scattered over a large area.

Note: You have to differentiate between the cluster sampling and the stratified (in the stratified sampling we choose students randomly even from each section, but in the cluster we take the whole section).

TABLE 10–1. Probability Sampling Chart

Type of Sampling	Description of Methodology	Advantages	Disadvantages
A. Simple random	Assign a number to each member of the population. Select the sample through a table of random numbers.	1. Little knowledge of population is needed 2. Most unbiased of probability methods 3. Easy to analyze data and compute errors	1. A complete listing of population is necessary 2. Time consuming 3. Expensive
B. Stratified	Divide population into strata. Determine number of cases desired in each stratum. Random sample these subgroups.	1. Increases probability of sample being representative 2. Assures adequate number of cases for subgroups	1. Requires accurate knowledge of population 2. May be costly to prepare stratified lists 3. Statistics more complicated
1. Proportionate	Determine sampling fraction for each stratum that is equal to its proportion in the total population.		
2. Disproportionate	Sample is drawn in manner to ensure that each stratum is well represented. Used when strata are very unequal.		
C. Cluster	Groups rather than people are selected from population. Successive steps of selection are done (state, city, county). Then sample is randomly selected from clusters.	1. Saves time and money 2. Arrangements made with small number of sampling units 3. Characteristics of clusters as well as those of population can be estimated	1. Larger sampling errors than other probability samples 2. Requires assignment of each member of population uniquely to a cluster 3. Statistics are more complicated
D. Systematic	Obtain listing of population Determine sample size. Determine sampling interval ($k = N/n$). Select random starting point. Select every kth element.	1. Easy to draw sample 2. Economical 3. Time-saving technique	1. Samples may be biased if ordering of population is not random 2. After the first sampling element is chosen, population members no longer have equal chance of being chosen

Good Luck!