Biostatistics

Doctor 2017 | Medicine | JU

Number >>	3
Doctor	Dr. Hamza Alduraidi
Done By	Dana Alqatawneh
CorrectedBy	Sarah Jaar

Quick review

The two main types of sampling are Probability Sampling and Non-Probability Sampling. In probability sampling, every member of the population has an equal chance of being selected. The population is the sampling frame. The types of Probability sampling are *simple random sampling*, *systematic sampling*, *stratified sampling*, *multistage sampling*, and *cluster sampling.* Probability sampling is ideal but difficult to perform.

In this sheet, Non-probability sampling, types of variables, and levels of measurement are discussed.

Non-Probability Sampling

Sometimes the circumstances, the nature of the population to be studied on, or the nature of the research question make the researchers uninterested in or unable to guarantee an equal chance for every member of the population to be selected. So, they use non-probability sampling. Non-probability sampling is when the sample elements are chosen from a population by nonrandom methods. Non-probability sampling helps answer the research question, and helps attain a sample suitable for testing the hypothesis even with all the limitations. Nonprobability sampling is more likely to produce a biased sample than probability sampling. This restricts the generalization of the study findings. The most frequent reasons for use of nonprobability samples involve convenience and desire to use available subjects.

Ex. A study conducted on diabetic patients in Jordan to measure their adherence to medication (if they tend to skip pills). There are many limitations to perform probability sampling in this study; we do not have the list of all the names of diabetic patients in the country, and there is no access to all patients. A possible design for this study is to go to a diabetes clinic several times throughout the day and assign patients available and willing to participate in the study. There is no equal chance for all diabetic patients of participating in the study.

Ex. Assume that 10% of the population are women, and 90% are men. A conducted study is interested in studying both genders equally (equal chance is not a concern). The subjects enrolled in the study would be 50% women and 50% men.

Types of nonprobability sampling methods:

1. *Convenience sampling*.
2. *Snowball sampling*.

3. *Quota sampling*.
4. *Purposive sampling*.

1. *Convenience sampling.* (accidental or incidental sampling)

This method is the easiest, and simplest way of sampling. It saves time and money, and it is the most frequently used method in health research studies. People may or may not be typical of the population. There is no accurate way to determine their representativeness.

> Ex. A study that requires subjects to fill a questionnaire. Every 2 hours, data collectors visit a waiting room in a hospital which has 20 people per hour; they ask patients to fill up the questionnaire only if they wish to. If they visited 5 times they would have had a sample of 100 subjects. (convenient method)
> Ex. Putting up a flier calling for anyone interested in joining the research study. (convenient to the researchers)

The disadvantages of this method are that it is non-ideal, non-representative, and it is subject to lots of sampling bias. However, it is the only available option due to time, limited resources, limited budget, and limited knowledge of the population. It also depends on the nature of the research question.

2. *Snowball sampling*.

This method is when study subjects assist in obtaining other potential subjects (networking). It is called snowball sampling because at first the sample is small then it gets bigger and bigger like a rolling snowball. It is used in topics of research where the subjects are reluctant to make their identity known, such as drug users and AIDS patients, and in rare diseases such as certain types of lymphoma. These populations are difficult to find especially if there is no organization that provides care for patients of such diseases.

> Ex. A study on mothers of autistic children. They are difficult to find but as soon as one mother interested in the study is found, she can give the researchers contact information of other mothers. These mothers can give contact information of others and so on and so forth.

3. *Quota sampling.*

This method is when a percentage of subjects in the study is given to a group of people that are a minority in the population. In *quota sampling,* variables of interest to the researcher (include subject attributes) such as age, gender, ethnicity, religion, and educational background are included in the sample. *Quota sampling* is a method for securing potential subjects from these strata. This minority group would be overrepresented in the study in comparison to their percentage in the population. (This is similar to the quota given for women in the parliament where a certain percentage of seats will definitely be filled by women even if they attained less votes than men that did not win a seat.)

Quota sampling is similar to *stratified sampling* in that the first step involves dividing the population into homogeneous strata and selecting sample elements from each of these strata. In *stratified random sampling* the sample is done by random selection (a percentage of each strata is selected randomly). In *quota sampling*, the sample is selected by convenience (ex. the first 50% of males and 50% of females).

> Ex. A study involves different races. Assume that in a population there are four races: Black (15%), White (15%), Latino (10%) and Native (60%). In *stratified sampling*, the percentages are respected meaning that the sample in the study will have 15% of the black race, 15% of the white race, 10% of the Latino race, and 60% of the Native race. However, in *quota sampling* the minority groups are overrepresented, meaning that the sample of subjects will have 25% of each race. The proportions in real life are not favorable.

4. *Purposive sampling.* (handpicking, judgmental)

Subjects are chosen because they are typical or representative of the accessible population, or because they are experts (more knowledgeable) in the field of research topic. Equal chance for all the population is not favorable in this case. When *purposive sampling* is used, the results are generalized on the total population regardless to the selectiveness of the sample. Qualitative researchers use *purposive sampling.*

> *Ex.* Any study about a political topic. Subjects in the study would be from political activists that have an idea/ opinion about the topic. It is unwise to include subjects from the population that have no idea about the topic.

Purposive sampling is very similar to *quota sampling.* Do not confuse them! In *purposive sampling,* you go to a demographic (ex. the educated) and ignore all other demographic regions. While in *quota sampling,* all demographics are included but the percentages are manipulated to be more inclusive.

You can combine the above methods of sampling depending on the study. For instance, in the example of the different races; you could use *quota sampling* for the Latinos, Blacks, and whites; then use *convenience sampling* for the Native population.

Note: In the exam, you will be given a scenario and you need to be able to distinguish and identify between the different types of probability and non-probability sampling. Focus! There might more than one answer correct, but don't worry it won't be complicated. Best of luck!!!

Variables

A variable is an object, characteristic, or a property that can have different values (it varies). The opposite of variable is a constant; it does not have different values.

Variables can be classified into independent and dependent variable. The dependent variable is the presumed outcome or effect (of an independent variable); the independent variable is the presumed cause (of a dependent variable).

> Ex. the relationship between smoking and lung cancer. Smoking is the independent variable, and lung cancer is the dependent variable.

Variables could be quantitative in which they can be measured in some way, or they could be qualitative in which it is characterized by its inability to be measured but it can be sorted into categories.

From the slides:

Variables

represent information that must be collected in order to meet the objectives of a study

Measurable characteristic of a person, object or phenomenon which can take on different values: weight, distance, monthly income, number of children, color, outcome of disease, types of food, sex.

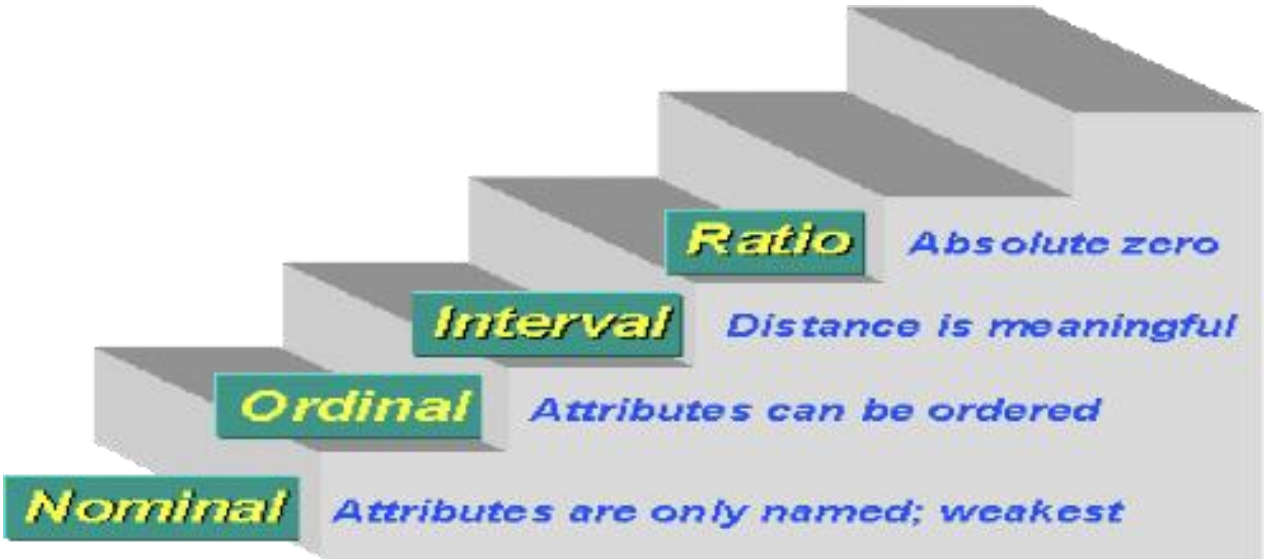
Variables allow clear definition of the core problem and influencing factors by introducing the concept of value.

Random variables: cannot be predicted in advance because they arise by chance. Observations or measurements are used to obtain the value of a random variable. Random variables are either discrete or continuous.

1. Discrete means individually separate and distinct. A discrete random variable has gaps or interruptions in the values it can have. The values may be whole numbers or have spaces between them.
   Ex. gender is a discrete variable; subjects could either be female or male.

2. Continuous random variables do not have gaps in the values it can assume. Its properties are like the real numbers. Fractions are accepted.
   Ex. age is a continuous variable. Subjects could be 19, 60, 45.25,35.5, …
   Height is also a continuous variable.

Levels of measurement



Discrete variables have 2 types: *Nominal* and *Ordinal*.

1. *Nominal discrete variables*
   It is the lowest level of measurement. Its categories are distinct from each other such as race, gender, religion, marital status, blood type. Numbers are given as

codes to be able to enter information in your computer or database for data analysis; they have no value or order, just symbols.
Ex. Giving females the number 1 and males the number 0. Subjects are either 1 or 0. No in between. The results won't be affected if you reversed the numbers; females 0 and males 1.

When there are only 2 options for a variable (nominal variable), it is called dichotomous.

Appropriate statistics used: mode and frequency. Average cannot be used; it is meaningless. (to be discussed in detail in a separate lecture.)

2. *Ordinal discrete variable.* (ordinal=order) count and percent
   Involves using numbers to designate ordering on an attribute. The exact differences between the ranks cannot be specified such as it indicates order rather than exact quantity. Numbers have mathematical value.
   Ex. A research question involves educational level. The greater the no. the higher the educational level.
   1: high school diploma
   2: bachelor's degree
   3: master's degree
   4: PhD degree
   Ex. age groups could be divided into childhood, youth, adulthood, elderly. They are numbered like the previous example.
   Ex. anxiety level: mild, moderate, severe. They are numbered.

The appropriate statistics used is mode, frequency, median, but not mean. (This will be discussed in a separate lecture.)

Continuous variables have 2 types: interval and ratio. Mean and standard deviation.

1. Ratio (most accurate and significant; most valid and reliable)
   This level of measurement is the highest level of data where data can be categorized, ranked. The difference between ranks can be specified and a true or natural zero point can be identified. A zero point means that there is a total absence of the quantity being measured. All scales start at 0; such as mass in kilograms or pounds. 0 has a meaning and is not arbitrary (subjective).
   0 means null.
   Ex. total amount of money.
   Do not confuse this with interval level of measurement

2. Interval

   They are real numbers and the difference between the ranks can be specified. 0 does not mean absence of, it's just another number (no "true zero"). They involve assigning numbers that indicate both the ordering on an attribute, and the distance between score values on the attribute. They are actual numbers on a scale of measurement. It has equal intervals (explained in example)

   Ex. body temperature on Celsius thermometer (36.2,37.2) There is a difference of 1.0 degree in body temperature. 0 degrees Celsius is just like any other number.

Appropriate statistics: mode, frequency, mean, median.

It is important to know the type of dependent variable because it is crucial to determine which type of inferential statistics to use in further steps in the research project. For instance, if the dependent variable is continuous, a t-test is used. If the dependent variable is dichotomous, ordinal, or nominal, a t-test cannot be used, instead a chi squared test is used.

Best of luck.