Lecture 1. Reading
Author: Bondor Cosmina

Variables

When the doctor consults the patient, he notes in the file various data about the patient and his condition, about the symptoms reported by him, symptoms that are not necessarily present, but may appear and disappear over the past few months. When the doctor is only interested in this patient, he will write down the respective data in the file or not, depending on the various criteria he has regarding this patient. For example, he will be interested in the diagnosis that the patient obtained and the medication he prescribed, but he will be less interested in the symptoms reported by the patient on the basis of which he made these diagnoses. Thus, patient files do not contain all the data necessary to recreate the steps necessary for their diagnosis. If someone is interested in research and wants to observe phenomena such as risk factors for certain diseases, he will need data from several patients. Such answers cannot be based on a single example.

Example. A client of an obstetrics and gynecology clinic is discharged after a successful birth. The gynecologist warns her to try not to get pregnant for the next 6 months – a year, because statistics show that miscarriages are more common in the year after giving birth. The client replies, how so? My cousin got pregnant 5 months after her first birth and did not have a miscarriage, but carried the pregnancy to term. The doctor points out the difference between studying a single case and studying enough cases. But the patient does not seem to be convinced by this and a few months later she presents with a miscarriage. In conclusion, when we are interested in researching a medical phenomenon we will study several individuals, that is, we will repeat the measurements/observations or experiments in the case of several people. This repetition will help us organize ourselves in order to draw conclusions. An ordering, classifications will be needed to observe the laws that exist in nature and their repeatability.

Variability

Some characteristics of patients can be very different depending on the individual. For example, weight measured in kg. If we have a scale that measures body weight with 4-5 decimal places, then we will hardly find among the 2000 patients of the family doctor two with the same weight. But other characteristics are not so different between individuals. For example, the flu test performed on the patient was positive, negative or invalid. There are no other possibilities in the case of the test for diagnosing flu.

Depending on this variability of a characteristic, it was classified.

Variable Types

In research, variables are classified according to their type. It is important not to confuse variables with the values that they can take. For example, body weight is the name of the variable, 34 kg is one of the possible values of the variable, a value that is also called the data. Another example is the variable named Obese. Overweight is a possible category of this variable, in this case we cannot talk about values of this variable.

So, depending on the values or categories that a variable can take, we will classify variables into nominal, dichotomous or ordinal qualitative variables, continuous or discrete quantitative variables and survival variables.

Quantitative variables

A quantitative variable has possible values, these can be measured using measuring instruments or observed by counting. The possible values are numbers. With these numbers, mathematical operations such as summation, subtraction, averages, etc. can be performed.

Continuous quantitative variable

A continuous quantitative variable has as values the set of real numbers or value intervals. It is called continuous because the measurement space is continuous, that is, it can be measured with any number of decimal places and these make sense. Let's take body weight as an example. Between 34 and 35 kg there is an infinity of values, all possible when measuring body weight. Even if the weight will be noted with only two decimal places or even without decimals, this variable is a continuous variable. Because in research we need to understand the data of several patients, continuous variables are accompanied by the units of measurement. The units of measurement will be specified both in the data collection forms and in the tables or figures resulting from the research. Other examples of continuous variables: the length of a tooth's root (mm), saliva flow rate (ml/min), bite force (N), tooth eruption age (years), etc.
Discrete quantitative variable
A discrete quantitative variable has as values the set of whole numbers, that is, numbers without decimals. The measurement space is not continuous; decimals have no meaning. Let's take for example the number of teeth in a patient's oral cavity. We will not say that he has 7.3 teeth, decimals have no meaning in this case. Usually these variables are not accompanied by units of measurement. Other examples of discrete variables: number of cavities (decayed teeth), number of dental visits per year, number of extracted teeth, number of teeth brushes per day etc.

Qualitative variables

A qualitative variable has possible categories. Categories do not have numerical properties.

Dichotomous qualitative variable

A dichotomous qualitative variable has two categories. A dichotomous variable is a nominal variable, but with a limited number of categories. Sometimes it is possible to have more than one category, but the data collected only has two categories. The variable will be considered a dichotomous variable. We are interested in the type of variable, because we are going to analyze it to draw various conclusions of medical interest using statistical procedures, so we are interested in the type of variable. Let's take for example tooth sensitivity with two possible values: presence, absence. Other examples of dichotomous variables: Smoking (Yes/No), orthodontic treatment (Present/Absent), Caries (Present/Absent), gum disease (Present/Absent), etc.

Nominal qualitative variable

A nominal qualitative variable has more than two categories. There is no possible order between the categories. Let's take for example tooth sensitivity with three possible values: presence, absence, absence. Here, missing refers to a date that was not collected or was not found in the patient's record. Other examples of nominal variables: smoking (never smoking, former smoker, current smoker), dental interventions (scaling, professional brushing, air-flow), dental appliances (fixed metal brackets/arch, fixed with sapphire brackets/arch, lingual dentures, etc.), etc.

Ordinal qualitative variable (ordered)

An ordinal qualitative variable has categories that can be ordered in some order. For example, Oral hygiene index (Good, Fair, Poor). Ordinal variables in dentistry facilitate the classification

of different medical conditions, since the distances between categories are not quantifiable or the same.
An ordinal variable can be coded with numbers, as in the case of a score. The distance between categories is not the same as in the case of a discrete variable. For example, the number of children in a family is a discrete variable, but the grade in English conversation is not a discrete variable, even if it is numerical. The student who gets a grade of 10 does not know conversation twice as well as the student who got a grade of 5, but only better. The distance between a grade of 9 and a grade of 10 is not the same as the distance between a grade of 4 and a grade of 5. So such a variable does not have numerical significance, even if it was noted with numbers. Similarly, grades could be given in categories such as: very good, good, etc. This observation can also be valid in classifying any score. For example, a "Gingival Index" score of 0, 1, 2, or 3 would be ordinal, as it indicates different levels of gum health, with 3 being the most severe. But for example, an IQ score based on the number of exercises answered correctly, with possible scores between 0 and 250 and above, is a score that can be considered numerical and treated as such from a statistical point of view. In the future, we will classify variables according to what can be processed against them. We will need to make a decision for each score separately. It will be presented as a mean or median. If the data leans towards a numerical significance, then we will present the mean, especially if we are talking about a score such as an IQ score with many possible values. If the data leans towards a non-numerical significance, then we will present the median or frequency table for each category separately.
Other examples of ordered variables: Dental Anxiety Level (low, moderate, high), Stage of Periodontal Disease (Stage 1 –Stage 4), plaque index (0-3), tooth sensitivity score, pain score (0-10), gingival index score, etc.

Survival Variables

There are studies in which time is the variable we want to study. This type of variable is known as the survival variable and is the time to a certain event. For example, time to survive, time to heal, time to a substance disappearing from the body.

About variables - uses

Usually, in studies more than one variable is recorded per individual. The purpose of a study can be to determine if and/or how one or more variables affect another.

Response Variable

The outcome of a study. A variable you would be interested in predicting. Often called a dependent variable or predicted variable.

Explanatory Variable

Any variable that explains the response variable. Often called an independent variable or predictor variable.

Calculated variables

Some variables can be calculated based on other variables using formulas.
There are also variables that can be reduced from a continuous variable to an ordinal or dichotomous variable. For example, obesity can be collected as the body mass index (kg/m2) of the person or as an ordinal variable with categories normal/overweight, obese or as a dichotomous variable Presence/Absence. By reducing the variable to a qualitative form, what we can state or observe about the variables is also reduced. Even though the medical understanding of a

phenomenon is greatly simplified by a reduced form of a variable, it is desirable when collecting to collect the variables in their continuous forms if possible. These can be reduced later, but the reverse process is not possible.

Basic statistic vocabulary in research methodology

Population

A population is any entire collection of people, animals, plants or things from which we may collect data. It is the entire group we are interested in, which we wish to describe or draw conclusions about.

In order to make any generalisations about a population, a sample, that is meant to be representative of the population, is often studied.

It is important that the investigator defines the population before collecting the sample.

Sample

A sample is a group of units selected from a larger group (the population). By studying the sample, it is hoped to draw valid conclusions about the larger group.

The sample should be representative of the general population. This is often best achieved by random sampling.

Statistic

A statistic is a quantity that is calculated from a sample of data. It is used to give information about unknown values in the corresponding population. For example, the average of the data in a sample is used to give information about the overall average in the population from which that sample was drawn.

Parameter

A parameter is a value, usually unknown (and which therefore has to be estimated), used to represent a certain population characteristic. For example, the population mean is a parameter that is often used to indicate the average value of a quantity.

Within a population, a parameter is a fixed value which does not vary. Each sample drawn from the population has its own value of any statistic that is used to estimate this parameter. For example, the mean of the data in a sample is used to give information about the overall mean in the population from which that sample was drawn.

Parameters are often assigned Greek letters, whereas statistics are assigned Roman letters.

Statistical Inference

Statistical Inference makes use of information from a sample to draw conclusions (inferences) about the population from which the sample was taken.

A sample statistic gives information about a corresponding population parameter. For example, the sample mean for a set of data would give information about the overall population mean.

Bibliography

1. Tudor C. DRUGAN, Sorana D. BOLBOACĂ, Daniel LEUCUTA, Cosmina BONDOR, Tudor CALINICI, Madalina VALEANU, Horatiu COLOSI, Mihaela IANCU, Dan ISTRATE , Curs de Biostatistica Medicala, editia a 2-a, Editura Medicala Universitara "Iuliu Hatieganu" Cluj-Napoca, 169 pgs, 2018, ISBN 978-973-693-861-0
2. Rosner, Bernard (Bernard A.). Fundamentals of Biostatistics. Boston :Brooks/Cole, Cengage Learning, 2011.
3. Robert H. Riffenburgh, Daniel L. Gillen. Statistics in Medicine: 4th ed. Oxford, UK: Academic Press, 2020.

4. Lang TA, Altman DG. Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or The SAMPL Guidelines" in: Smart P, Maisonneuve H, Polderman A (eds). Science Editors' Handbook, European Association of Science Editors, 2013. Available from: https://www.equator-network.org/wp-content/uploads/2013/03/SAMPL-Guidelines-3-13-13.pdf (accessed 19 October 2024)