

Study variables in scientific research

Dr. Gerardo García Maldonado
Full-Time Professor and Researcher
SNII Level 1
Autonomous University of Tamaulipas
Faculty of Medicine of Tampico “Dr. Alberto Romo Caballero”
gmaldonado@docentes.uat.edu.mx

Dr. José Eugenio Guerra Cárdenas
Full-Time Professor and Researcher
Autonomous University of Tamaulipas
Faculty of Medicine of Tampico “Dr. Alberto Romo Caballero”
jguerra@docentes.uat.edu.mx

Dra. Ma. Elena Calles Santoyo
Part-Time Professor and Researcher
Autonomous University of Tamaulipas
Faculty of Medicine of Tampico “Dr. Alberto Romo Caballero”
mecalles@docentes.uat.edu.mx

INTRODUCTION

In scientific research, regardless of the field of knowledge, the so-called study variables represent one of the fundamental pillars of the research process. That is, they are elements that guide the methodological and statistical development of a project. Understanding this circumstance—knowing what variables imply and represent both conceptually and operationally, that is, how they will be measured or quantified—is crucial to ultimately being able to analyze and interpret results. Evidently, depending on the discipline, study variables will be specific. In other words, variables in Chemistry, Physics, or Agronomy have their own characteristics compared to those in the medical field.

Undoubtedly, in the latter field, variables include diseases, treatments, diagnoses, and all aspects related to the medical profession, as well as everything inherent to human beings in their personal, social, and family dimensions, which constitute elements of interest for medical practice. Research variables are the characteristics, conditions, or attributes of the elements that make up a given field of knowledge and that may change or vary within a study. Their distinctive feature is that they can be measured, observed, or manipulated to answer a research question or test a hypothesis.

Each variable plays a specific role in the research process, helping to describe, analyze, and understand the phenomena under study. Always guided by the principle of parsimony, a taxonomic organization of research variables is presented below. Different types were included in a single figure based on their nature, measurement scale, and participation in statistical processes.

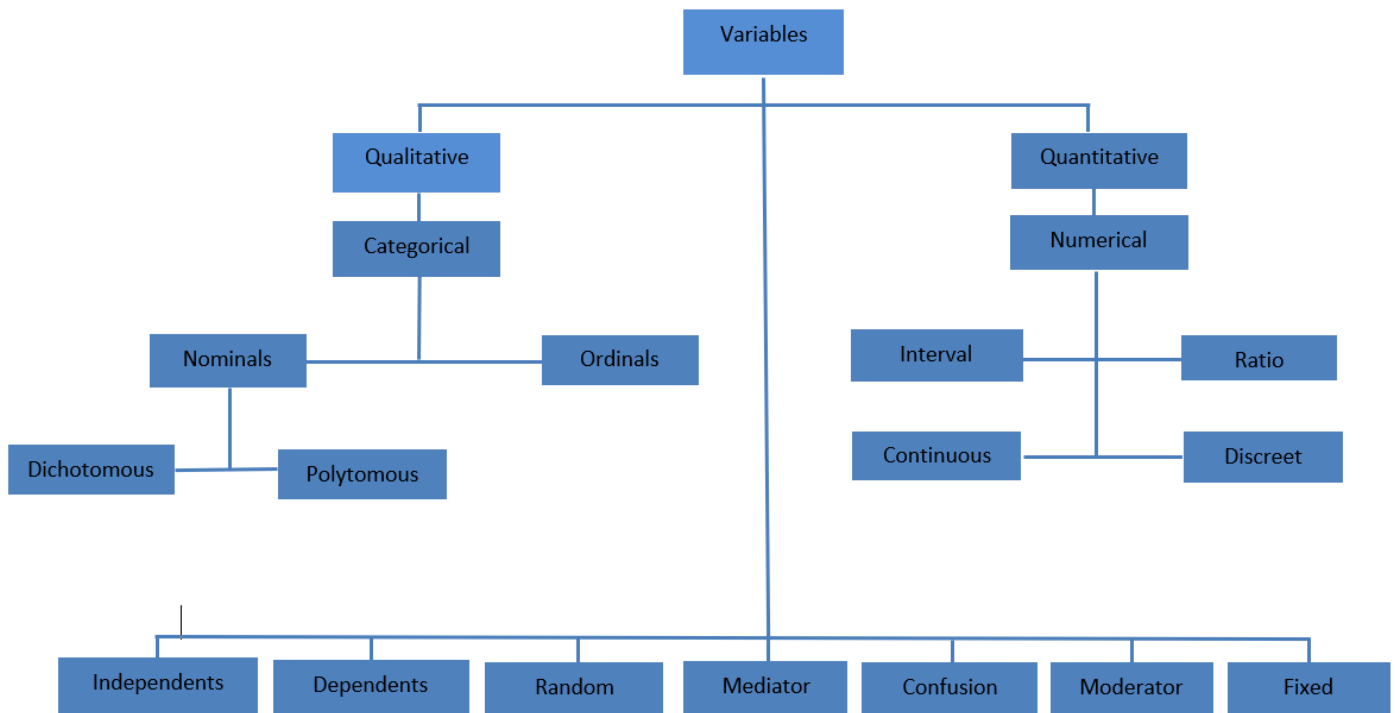


Figure 1. Variables in scientific research

Before addressing each element of this taxonomy, it is worth considering some preliminary aspects.

OBJECTIVITY AND SUBJECTIVITY OF VARIABLES

First, it is important to understand that variables can be objective or subjective. But what are the differences?

Fundamentally, the difference lies in the way they are measured and in the nature of the data they represent. However, other aspects must also be considered, as shown in the following table.

Table 1. Differences between objective and subjective variables

<p>OBJECTIVES</p>	<ul style="list-style-type: none"> • They are quantified directly, without interpretation, and are independent of the observer’s opinion. • They generally allow precise measurements. <p>Example:</p> <ul style="list-style-type: none"> • Body weight (kg), blood glucose level (mg/dL), height (cm). • The measurement procedure is standardized; therefore, different observers are expected to obtain the same results.
<p>SUBJECTIVES</p>	<ul style="list-style-type: none"> • They depend on the perception of the observer measuring them. • They cannot be quantified precisely. • They may be influenced by beliefs or incorrect interpretations. <p>Example:</p> <ul style="list-style-type: none"> • Perceived quality of medical care, quality of life perception, pain intensity, satisfaction level, degree of depression. • These variables are often measured through interviews or surveys.

Dicho de otra manera, las variables objetivas se refieren a las que por la naturaleza de sus datos tienen propiedades medibles (ejemplo: glucosa en sangre, peso corporal), es decir, representan condiciones físicas. En el caso de las variables subjetivas, estas representan una idea o concepto (percepción de bienestar, depresión) que pueden tener diferentes enfoques, por lo que es fundamental desglosar el constructo en forma estructurada y racional, para conformar conceptos y dimensiones que puedan ser medibles y determinar si dicho constructo está presente o no en la evaluación.

DIMENSIONALIDAD DE UNA VARIABLE

In scientific research, the dimension of a variable refers to the nature or specific characteristic of the variable that is being measured or studied. We could preliminarily say, for example: If the variable body mass index is studied, the dimensions would be (weight and height). If the variable quality of life is studied, the dimensions could be (physical health, mental health, personal satisfaction, family relationships, for example).

By identifying the dimensionality of a variable, researchers can:

- Define precisely what is being studied
- Select appropriate measurement instruments
- Analyze and interpret data accurately
- Obtain more specific and relevant results

Objective variables are therefore composed of physical dimensions, and subjective variables of analytical dimensions, also called logical dimensions. Although the distinction is made (logical dimension–physical dimension), which is reasonable in essence, in practice the term dimensionality is used more generally, since although they are semantically different, in the context of scientific research, when talking about dimensionality it is understood that it can be logical or physical. We can also add that a variable may have one or several dimensions, and would therefore be described as unidimensional or multidimensional. Understanding the considerations of dimensionality, objectivity, or subjectivity, we can then establish the following considerations:

Subjective, Logical, Multidimensional Variable:

For example:

QUALITY OF LIFE. - It is composed of several dimensions: physical health, personal satisfaction, family relationships, and environment.

DEPRESSION. - Affective, motor, cognitive, and behavioral dimensions are integrated.

As can be observed, each of these dimensions captures a different aspect, which together make it possible to understand what each construct refers to. With this, we understand that it would be possible to speak of quality of life in terms of family relationships, but not necessarily physical health, or to consider depression in terms of affectivity, but not cognitive or motor aspects. It would be difficult to measure the variable quality of life or depression as something unidimensional; therefore, the need to break down the construct into different components is emphasized for better understanding and for more specific statistical calculations. Finally, it is worth reiterating that these are conditions that are not measured physically.

Objective, Physical, Multidimensional Variable:

They are composed of two or more dimensions, but their difference is physical; therefore, they are repeatable, reproducible, and have precise measurement.

For example:

BODY MASS INDEX. - Weight and Height.

BLOOD PRESSURE. - Systolic and Diastolic

RESPIRATION. - Inhalation and Exhalation

RESPUESTA INMUNE. - Cellular and Molecular

VELOCITY. - Distance and Time

DENSITY. - Mass and Volume

HEAT FLOW. - Temperature and Surface Area

On the other hand, if the variables present a single dimension and are likewise integrated by their objectivity or subjectivity, we would have:

Objective, Physical, Unidimensional Variable:

For example: Heart rate, respiratory rate, cardiac enzymes (Troponin, Creatine Kinase), body temperature, to name a few.

These are therefore individual physical conditions with a single measurement and dimension, which are quantified directly.

Subjective, Logical, Unidimensional Variable:

For example:

Pain Level.

Unidimensional subjective variables are not very common in research; likewise, as with multidimensional subjective variables, they hardly achieve stability under the criteria of repeatability and reproducibility. In other words, they involve a single dimension and measurement and depend on the evaluator's perception.

MEASUREMENT RESOURCES

The instruments used to measure variables are diverse and depend on their nature, what is to be measured, and the researcher's selection. For a unidimensional objective variable, evidently only one instrument or resource will be required, often mechanical:

For example:

For body weight, a scale; for height, a stadiometer (also called a height meter); for body temperature, a thermometer, etc.

If the variable is objective and multidimensional:

For example:

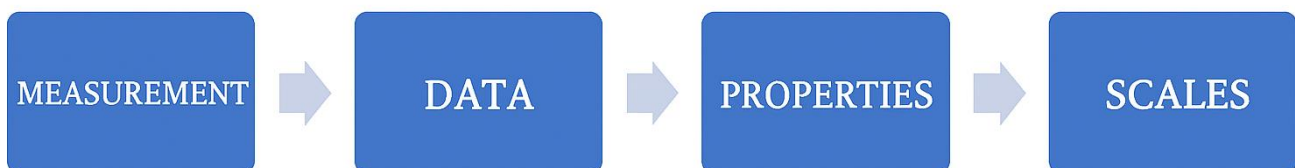
For body mass index, the combination of dimensions will therefore be required and, consequently, of resources or instruments: $(\text{weight}/\text{height}^2)$ (scale and stadiometer). Other examples would include mass, density, immune response, etc.

In the case of unidimensional or multidimensional subjective logical or analytical variables:

The resources are called documentary, such as scales, questionnaires, or inventories (both unidimensional and multidimensional); that is, no physical devices are used.

In recent years, a wide variety of these instruments have been developed and validated for different purposes. Ideally, once the study variable has been decided, its dimensionality and the best resource for its measurement must be established. It is important to note that measuring a variable is only the beginning of a series of considerations that must be taken into account for the purposes of scientific research.

The following diagram illustrates what has been described.



- By measuring, we obtain data
- The final values of the data have properties
- The properties make it possible to determine the measurement scale to be used

* Do not confuse measurement scale with measurement resource.

The final values of the data obtained from measuring the variable can generally be categorized as categorical or numerical (the former reflect categories or characteristics, while the latter come from counting or measuring). The

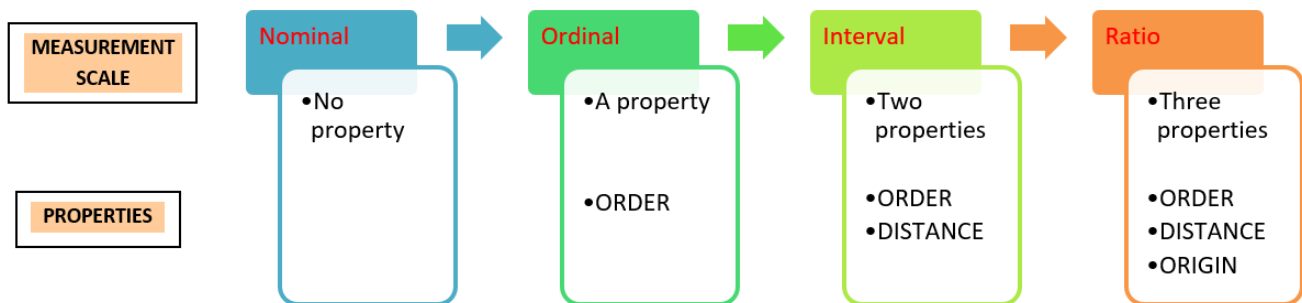
former are also known as qualitative and the latter as quantitative, although this terminology is infrequently used in this context because it leads to confusion.

The properties of the data can be classified according to the following conditions:

- Order
- Distance
- Origin

Consequently, depending on these properties, the measurement scales are configured (nominal, ordinal, interval, and ratio).

In this way, the following scheme is established:



Even though it is common to state, quoting literally, “nominal variable, interval variable, ordinal variable, etc.”, in a strict sense it should be stated as: “measurement scale of the ordinal variable, measurement scale of the interval variable, etc.” According to the taxonomy shown in Figure 1, qualitative variables are also designated as categorical based on their final value; these, in turn, encompass the nominal and ordinal measurement scales. Quantitative variables, for their part, correspond to the so-called numerical variables, which include the interval and ratio measurement scales, as well as the so-called discrete and continuous variables.

CATEGORICAL VARIABLES

NOMINAL SCALE

These are variables that correspond to categories; therefore, they do not include order, sequence, or hierarchy.

This type of variable can be grouped into two modalities:

Dichotomous: Refers to a division into two opposite and mutually exclusive parts.

For example:

Gender variable: male or female

Health status variable: healthy or ill

Diabetes variable: diabetes present or diabetes absent

Polytomous: Refers to something that is divided into many parts or elements, which allows a broader understanding of the variable under study.

For example:

Marital status variable: single, married, divorced, widowed, common-law

Religion variable: Catholic, Jewish, Muslim, Buddhist

ORDINAL SCALE

In this case, the categories have an order or hierarchy, but there is no way to quantify distances, since they do not include this property as numerical variables do.

For example:

Educational level variable: primary, secondary, high school, bachelor's degree

Depression intensity variable: mild, moderate, severe, profound

NUMERICAL VARIABLES

INTERVAL SCALE

The expression is numerical, so its main property is to possess, in addition to hierarchy, a quantifiable distance; therefore, measurement is objective. However, in this case, the zero point is arbitrary, that is, it does not represent the nonexistence of the variable or the absence of what is being measured; it is merely a conventionally chosen reference point. This type of variable allows differences between values to be measured, since they are constant, but not proportional relationships, precisely because the zero is arbitrary.

For example:

If we say that the temperature in degrees Celsius is zero, this does not mean that there is a total absence of temperature; zero represents the freezing point of water. Furthermore, there are negative values, such as -1 degree, -2 degrees, etc.

In this case, if we said that 10 degrees Celsius is half of 20 degrees, we would be making an incorrect proportional relationship precisely because the zero is arbitrary.

RATIO SCALE

This is the scale with the highest hierarchy, since in addition to order and distance, it includes the property of origin, and the intervals between values are constant. Any type of mathematical operation can be performed, as well as proportional comparisons. Thus, zero in this case can be considered absolute, that is, not arbitrary, which signifies the absence of the unit of study; therefore, it does not accept negative values.

For example:

In the case of weight or length, zero does indicate the total absence of the measured property, and proportional relationships are meaningful; for example, an object that weighs 10 kg weighs twice as much as one that weighs 5 kg, or an event that lasts 4 hours lasts twice as long as one that lasts 2 hours. On the other hand, no object or person can weigh zero kilograms.

It is important to note that within the category of numerical variables, in addition to those defined by the measurement scale, there are others that are not established from this perspective and depend on the values they can take and how they are obtained.

DISCRETE VARIABLES

These are variables that take specific values derived from counting, generally whole numbers; that is, they do not admit fractions or intermediate values between one number and another. Less frequently, they are referred to as discontinuous. The values are finite and represent separate quantities.

For example:

Number of children in a family, number of patients, number of census beds in a hospital, number of deaths:
(1, 2, 3, 4...)

CONTINUOUS VARIABLES

They can take any value within an interval, including fractions and decimals. These variables are measured, not counted, and their possible values can be infinite.

For example:

A person's height: 1.80 meters

Drug concentration in blood: 0.8 mcg/ml

Body temperature: 36.5 degrees Celsius

In Figure 1, mention is made of another type of variables which, according to their configuration and positioning, are more closely related to the development of statistical processes.

INDEPENDENT VARIABLE

This is a variable that is manipulated or selected to observe how it affects another variable called the dependent variable. That is, the first is the factor that influences the outcome of the second.

It has some characteristics:

- It is controlled and defined by the researcher, especially in experiments or clinical trials.
- It can be categorical or numerical.
- It is used to explain, predict, or influence the dependent variable.

DEPENDENT VARIABLE

Also called response, outcome, or criterion variable, it is the variable that is measured to evaluate the effect of the independent variable.

Characteristics:

- It is the result of changes in the independent variable.
- It is measured by the researcher but not manipulated.
- It can be categorical or numerical.

In a study on the impact of insomnia on depression:

Independent variable: insomnia

Dependent variable: depression

It is important to note that it is the researcher's decision to determine which variable will be independent and which will be dependent, according to the line of research. It could therefore be the opposite, that is, the effect of depression on the development of insomnia:

Independent variable: depression

Dependent variable: insomnia

RANDOM VARIABLE

This is a variable whose specific outcome is uncertain, but whose general behavior can be known, described, or predicted through a probability distribution. This means that this modality indicates how values are distributed and how likely each one is to occur.

Example:

If we roll a die, we do not know which number will appear, but each number has a probability of 1/6 of appearing. If we toss a coin, we do not know whether it will land heads or tails, but each side has a 50% probability of appearing.

FIXED VARIABLE

This is a variable whose value does not change during an experiment or procedure. Unlike random variables, its values are constant or predefined and are not subject to chance or uncertainty. Main characteristics:

- Determinism: constant or defined value
- Control: used as a reference value
- Non-random: not influenced by uncertain factors

They provide a stable basis for interpreting results.

Example:

If the effect of diet on body mass index is studied, the number of calories assigned to each study group (2,000 calories, for example) would be the fixed variable.

There are other types of variables that, due to the function they fulfill in the analysis, are also important.

CONFOUNDING VARIABLE

This is a variable that the researcher keeps constant or controls to prevent it from influencing the relationship between the independent and dependent variables. They are used to ensure that any change in the dependent variable is due exclusively to the manipulation of the independent variable.

Example:

In an experiment on the impact of sleep on academic performance, diet and the number of hours of study could be confounding variables to be controlled so that they do not affect the results.

INTERVENING OR MEDIATING VARIABLE

This is a variable that lies between the independent and dependent variables and explains the relationship between them.

Example:

If the effect of stress on academic performance is investigated, motivation may be a mediating variable, since stress can reduce motivation, which in turn decreases performance.

MODERATING VARIABLE

This is a variable that affects the direction or strength of the relationship between the independent and dependent variables.

Example:

If we examine the relationship between family support and well-being, participants' personality (extroverted/introverted) could be a moderating variable, since it could influence how strong the relationship is between family support and well-being.

Finally, let us imagine a project that investigates the impact of the number of hours of sleep (independent variable) on academic performance (dependent variable) in university students, taking into account motivation (mediating variable) and controlling confounding factors such as age and hours of study (control variables).

REFERENCES

- Andrade Ruiz HA. Bioestadística aplicada en ciencias de la salud: guía complementaria. 1ª ed. Guatemala: Fundación Gustavo Palma Calderón; 2019.
- Bland M. An Introduction to Medical Statistics. 4th ed. Oxford: Oxford University Press; 2015.
- Daniel W. Bioestadística: base para el análisis de las ciencias de la salud. 4ª ed. México: Limusa Wiley; 2000.
- Daniel WW, Cross CL. Biostatistics: A Foundation for Analysis in the Health Sciences. 10th ed. Hoboken (NJ): Wiley; 2013.
- González Betanzos F, Escoto Ponce de León MC. Estadística aplicada en psicología y ciencias de la salud. México: Manual Moderno; 2017
- Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken (NJ): Wiley; 2013.
- Montero Fernández J, Minuesa Abril C. Estadística básica para ciencias de la salud. 1ª ed. España: Universidad de Extremadura; 2018
- Rius Díaz F, Barón López FJ. Bioestadística. 1ª ed. España: Ediciones Paraninfo; 2005.
- Ruiz Muñoz D. Manual de estadística. 1ª ed; 2000