VARIABLES

A Matter of Kind and Degree

Jean-Frédéric Morin, Alessandra Bonci, Université Laval

Variables are measurable representations. As such, they are located at the interface between theoretical constructs and empirical observations. For example, the concept of 'social capital' is a theoretical construct (see **CONCEPT CONSTRUCTION**) which can be **OPERATIONALIZED** using the variable 'social club membership'. The variable 'social club membership' is easier to measure than the abstract notion of 'social capital'. It also has more conceptual meaning than the observation that person A is a member of a book society.

Deductive research identifies variables by operationalizing abstract concepts. Inductive research typically constructs variables from the observation of units (see **DEDUCTIVE, INDUCTIVE, AND RETRODUCTIVE REASONING**). However, irrespective of whether the research is deductive and theory-driven or inductive and empirically driven, variables occupy a central position in research methodology. They lie halfway between the theoretical and empirical realms.

One of the key features of variables is that they vary across **UNITS**. Any variable can have at least two distinct values (also called attributes). As Gary King, Robert Keohane, and Sidney Verba argue, 'a variable can represent anything whose values change over a set of units' (King, Keohane, and Verba 1994: 51). For example, the national economy is not a well-defined variable as it is unclear on which dimension it varies. Variables for a given economy include the main factor of production (with qualitative values, such as labour, capital, and land) and the annual growth rate (with numerical values, ranging from 0 to an infinite number).

This contribution is divided into three parts. The first part distinguishes dependent and independent variables, the second part introduces other types of variables, and the third part presents different types of values. The conclusion discusses the **EPISTEMOLOGICAL** assumptions underlying the notion of variables. The variables of 'social club membership' and 'economic growth' are used throughout for illustrative purposes.

DEPENDENT AND INDEPENDENT VARIABLES

In positivism, research aims to identify a causal relationship between two different types of variables (see **POSITIVISM, POST-POSITIVISM, AND SOCIAL SCIENCE**). The first type is the independent variable (IV), otherwise known as the explanatory variable, predictor variable, treatment variable, input variable, explanans, or X variable. The second type is the dependent variable (DV), also known as the explained variable, predicted variable, response variable, output variable, explanandum, or Y variable. The independent variable is a potential cause of the dependent variable. For example, we could argue that the more people are members of social clubs, the higher their life expectancy. In this case, 'social club membership' is the independent variable because it does not depend on life expectancy, and 'life expectancy' is the dependent variable because it depends on social club membership.

RESEARCH QUESTIONS typically focus on a dependent variable, rather than an independent variable. For this reason, research projects often examine only one dependent variable, although they may consider several different independent variables. In fact, entire research programmes, research communities and, by extension, **LITERATURE REVIEWS**, are often organized around a single dependent variable. For example, most researchers working on economic growth focus on the causes of economic growth, rather than its consequences. Therefore, the causal relationship between economic growth and economic inequality is more likely to be explored by researchers interested in understanding the causes of inequality, than by experts on economic growth.

HYPOTHESES connect independent variables to dependent variables. It is easier to test a causal relationship when the independent variable is genuinely independent (from the dependent variable)—i.e. when the variation of the dependent variable has no impact on the independent variable. For this reason, the independent variable should be logically prior to the dependent variable. In the **HYPOTHESIS** that social club membership increases life expectancy, the independent variable seems to be sufficiently independent from the dependent variable. Indeed, it is unlikely that longevity has a causal effect on social club membership at a given age.

It can be difficult to establish whether the independent variable is truly independent, even when it precedes the dependent variable on a temporal level. For example, we can assume that social club membership increases the likelihood of being a candidate in a future election. Social club membership may affect people's perception and encourage them to start a political career (Putnam 2001). However, someone with political ambitions may join different social clubs for strategic reasons, as a way of building useful connections. In this case, social club membership would be the dependent variable, rather than the independent variable. Both variables may actually reinforce each other in a positive feedback loop. This can create an **ENDOGENEITY** problem, making it difficult to measure the effect that one variable has on the other.

OTHER TYPES OF VARIABLES

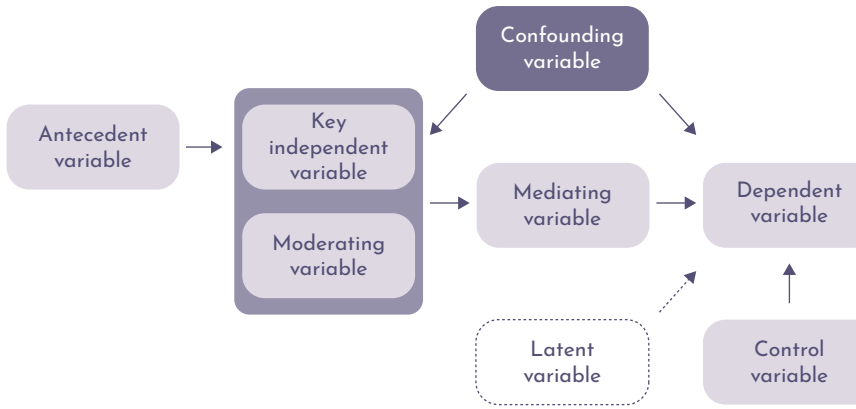
The interference of other types of variables can further complicate the identification of a causal relation between a dependent and independent variable. In social sciences, most phenomena are multicausal (see **MULTICAUSALITY AND EQUIFINALITY**). For example, economic growth is not caused by a single variable, but by several variables, including demographic growth, technological progress, rule of law, and trade openness. Measuring the impact of technological development on economic growth may be complicated by these other variables. Indeed, other variables may cause a high degree of variation on the dependent variable and, thus, conceal the significant relation between technological progress and economic growth.

This interference can be reduced by including control variables in the analysis. Control variables—also called extraneous variables—are not the primary focus of research. The researcher controls them in order to isolate the relationship between a dependent and independent variable. When a researcher claims that technological development leads to economic growth 'other things being equal', the control variables are these 'other things' that remain constant. For a methodological perspective, a variable can be controlled in numerous ways, including carefully selecting participants in **SURVEY RESEARCH** and **EXPERIMENTATION** (see **SAMPLING TECHNIQUES**), adding variables to the model in **REGRESSION ANALYSIS**, and comparing most similar cases in a **CASE STUDY** (see **CASE SELECTION**). Irrespective of the method used, a researcher should try to control all the variables that affect the dependent variables. Failure to include control variables may result in an omitted variable **BIAS**, which can lead to the overestimation or underestimation of the relation between the dependent and independent variables.

However, a perfectly controlled enquiry is extremely difficult to achieve in social sciences for at least three reasons. First, it is difficult to distinguish between relevant and irrelevant explanatory variables and make sure that all relevant control variables are considered. (Brady and Collier 2004: 59). Second, the number of **UNITS OF ANALYSIS** limits the number of variables that a researcher can control. Third, a researcher may find it challenging to document certain variables. In some cases, she or he might substitute a given variable for a proxy variable, which is strongly correlated with the missing variable. For example, gross domestic product (GDP) is frequently used as a proxy for population well-being. However, when no satisfactory proxy is available, the researcher may have to leave this variable latent, as illustrated in Figure 15.

One type of variable, which is important to control, is the confounding variable. This variable influences both dependent and independent variables. For example, economic downturns might lead to a political turnover, as well as to an increase in loans from the International Monetary Fund (IMF). If a researcher fails to account for the confounding variable of an economic downturn, he or she might find a spurious correlation between IMF loans and political turnovers—i.e. a covariation between two variables that have no causal link. If confounding variables are ignored, the null **HYPOTHESIS** may be rejected incorrectly and a correlation may be mistaken for **CAUSATION**.

FIGURE 15 Different types of variables



A moderating variable is another important type of variable. It can accentuate or diminish the relation between dependent and independent variables. For example, social club membership might increase annual income, but only for individuals who belong to the elite. On the contrary, an individual with a working-class background may not be able to use their social capital to generate additional income. In this case, social class is a moderating variable, which should be examined when studying the relation between social club membership and income.

Researchers can ignore other types of variables. For example, they may not consider antecedent variables (which precede the independent variables) and mediating variables, also known as intermediary variables (which transmit the effect of the independent variable to the dependent variable). Social club membership alone does not directly generate an increase in revenue. If we assume that there is a positive relation between the two variables, it is important to recognize that mediating variables also operate between the two. For example, the number of connections to potential clients is a mediating variable. However, it may be more difficult to determine the number of connections to potential clients than the number of social club memberships. A research design based on **PROCESS TRACING** can reveal the mediating variables. However, **REGRESSION ANALYSIS** often ignores variables, which are difficult to document or to substitute with a satisfactory proxy variable.

VALUES

When defining a variable, it is important to identify all of its possible values. The range of values should be exhaustive, in the sense that each observation should fit with at least one value, including 'unknown' and 'zero'.

There are different types of values. Categorical variables assign a nominal value to each observation, which is based on a given qualitative and non-hierarchical property. For example, the variable type of 'social clubs' can have the nominal values 'art clubs', 'sports clubs', and 'charity clubs'. Some analytical methods require that values be mutually exclusive, so

that one observation corresponds to a single value only (see **REGRESSION ANALYSIS**). Since one individual can belong to several different types of clubs, a researcher can dichotomize a categorical variable into a set of binary variables, known as dummy variables. Membership of an art club, a sports club, and a charity club are then considered to be three distinct variables, instead of three values of the same variable. There are only two possible values for each these dummy variable: yes (1) or no (0).

Other categorical variables have ordinal values. Unlike nominal values, ordinal values can be ranked in a specific order. For example, the prestige of a social club can be ranked into three values: low, intermediary, and high status. Increasing the number of ordinal values for a given variable may improve its precision, but often at the expense of reliability. Classifying the prestige of a social club in a ten-point scale can be difficult. In addition, the encoded value may not be consistent if the measure is repeated.

Variables can also have numerical values. The number of social club memberships is associated with a finite quantity of cardinal numbers, including a true zero point, which indicates the absence of social club membership. The variable 'average number of social club memberships per 1,000 inhabitants' is a ratio value: it has an infinite number of continuous values.

The types of value condition the type of analysis. Importantly, it is impossible to measure the arithmetic mean of categorical variables because the distance between different points of the scale is not necessarily equivalent. For example, although we could assess the prestige of thousands of social clubs on a scale from 1 to 10, their mean prestige score cannot be measured. The mode (the most common value) is the only central tendency measure that can be applied to nominal variables. The median (the value of the middle-ranked observation) can also be used for ordinal variables. Ratio scale is the most malleable type of value because it allows for several types of central tendency measures, such as arithmetic mean (Stevens 1946).

CONCLUSION

The language of variable and values is used primarily in the positivist tradition (see **POSITIVISM, POST-POSITIVISM, AND SOCIAL SCIENCE**). Several researchers from this tradition argue that the rules guiding quantitative analysis should equally apply to qualitative research (King, Keohane and Verba 1994). For example, they claim that the proscription of selecting cases based on the dependent variable (see **CASE SELECTION**) also applies to qualitative analysis.

However, some qualitative researchers reject the quantitative template and the variable terminology (Brady and Collier 2004). They argue that focusing on variables restricts their analysis unnecessarily. It forces researchers to explain phenomena, rather than describe or understand them. The variable terminology also suggests that relations between variables are mechanical and linear, despite the fact that they are often intertwined in complex unstable ways. Most critics of variable terminology prefer a more in-depth holistic approach, especially when the number of cases is small and **EPISTEMOLOGICAL** ambitions are idiographic.

REFERENCES

Brady, Henry and David Collier. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD, Rowman and Littlefield.

Hartmut, Esser. 1996. 'What is Wrong with "Variable Sociology"?' *European Sociological Review* 12(2): 159–166.

King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ, Princeton University Press.

Putnam, Robert D. 2001. *Bowling Alone: The Collapse and Revival of American Community*. New York, Simon & Schuster.

Stevens, Stanley Smith. 1946. 'On the Theory of Scales of Measurement'. *Science* 103(2684): 677–680.