

# Chapter 5

## Understanding Dependent and Independent Variables in Research Experiments and Hypothesis Testing



### 5.1 What Are Variables in Scientific Research?

*Variables* represent any quantifiable or measurable attributes in a given dataset. In theory, variable can be used to represent anything; ranging from some kind of phenomenon or entity one is trying to measure, to empirical study of events, ideas, subjects and objects or even time (Sarikas, 2020).

There are different ways in which variables are defined depending on the context it is used or applied.

In scientific research, “variable” can be defined as a measurable property or traits that changes (varies) or are affected as a result of the change across the experiment or hypothesis when testing. Perhaps, such properties or changes are captured regardless of whether the scientists or researchers are comparing the outcomes or relationships that exist among multiple groups of objects, multiple persons, or a single entity in an experimentation, e.g., performed over a period of time (Agravante, 2018).

In mathematics, the opposite of “variable” is referred to as “constant”. Therefore, we assume that a variable is an entity or quantifiable object with an *unknown value*, thus, the concept for which its application for scientific research is drawn.

Researchers often choose some kind of letters (annotations) to remind them of the quantities or parameters that are being measured or they are measuring (statistically represented). For example, in an equation, graph or linear regression model in which the “y”-axis or parameters can be defined as a function of “x”, i.e.,  $f(x)$ . It means that the value of  $y$  is dependent upon the value of  $x$ . Thus “x” and “y” values are described as the “variables”. In short definition, the term *variable* implies that the represented values or parameters can change over the course of a scientific experiment or hypothesis testing.

Logically, variables and the associated attributes can also be allied to the concept of necessary condition analysis (NCA) (Dul, 2016). Dul (2016) emphasized on the logic and methodology of “necessary-but-not-sufficient” conditions to define the

various processes or factors (e.g., measures of occurrences, features or characteristics of objects, etc.) that can be used for hypothesis testing, thus, leading to some sort of expected outcomes (conclusions). In essence, when defining the concept of NCA in relation to scientific research and experiments, the authors note that a necessary determinant (i.e., variables) must be present for achieving an outcome (or drawing conclusions). Although Dul (2016) states that the presence of the determinants (variables) is not always sufficient to prove that outcome, but instead to suggest or support the conclusions. In other words, variables for research purposes can only be used to draw causal inferences. More importantly, the validity of the resultant inferences or conclusion is grounded on the level of adequacy of the proposed or supported theories, quality of the measurement and analysis process, the statistical data analysis method or hypothesis testing, and SMART research design adopted by the researchers or scientists (Dul, 2016; Stone-Romero, 2002; Tynan et al., 2020).

Despite the fact that the necessary condition analysis (Tynan et al, 2020) is a theoretical method of its own in literature, it not only allows us to relatedly perform hypothesis testing through the underlying logic (indication) of *necessity of X for Y* in any statistical or rule-based analysis (e.g., IF – THEN statement or Association Rule Learning) (Okoye et al., 2014), but also the logic (NCA) can be grasped as an effective and efficient way to find relevant variables that must be leveraged or analyzed to support the occurring conclusions or hypothesis acceptance or rejection. Thus, variables are normally, in theory, represented as a group of items with several attributes or characteristics that can be combined (e.g., by adding or averaging) in order to obtain a concluding score for the said variables of interest or scrutiny (Dul, 2016).

Likewise, Tynan et al. (2020) note that although the NCA logic (in carrying out research experiments and hypothesis testing) enables the researchers to determine if the observed relations or association between the variables (e.g., independent versus dependent) (see: Sect. 5.1.1.1 and 5.1.1.2) are consistent with the “necessary-but-not-sufficient relation or condition” rule. It states that such condition suggests that the results of the experimentations or hypothesis testing are only probable—but not guaranteed—given the high levels or reliability of the considered variables (Tynan et al., 2020). Moreover, the outcomes may sometimes consequently lead to re-consideration (re-examination) of other variables that may have been apriori dismissed or considered as being irrelevant in the analysis process.

Theoretically, there are different types of variables, considering the diverse settings or context in which it is being used or applied. We note some of the examples to include: independent and dependent variables, intervening and moderator variables, constant or controllable variables, extraneous or predictor variables, etc. (Agravante, 2018). For all intents and purposes, this book focuses on the most frequently named or used variable when conducting the scientific research and data analytics as follows:

- Independent variables, and
- Dependent variables.

### 5.1.1 Types of Variables in Scientific Research

Deciding if a variable influences other variable(s) is one of the main challenges or components of conducting some scientific research particularly the social science. This is owing to the fact that by establishing the effects or assertions from the examined variable(s), the investigators are able to accept or reject a hypothesis, and, in turn, create a new knowledge about the phenomenon that is being studied.

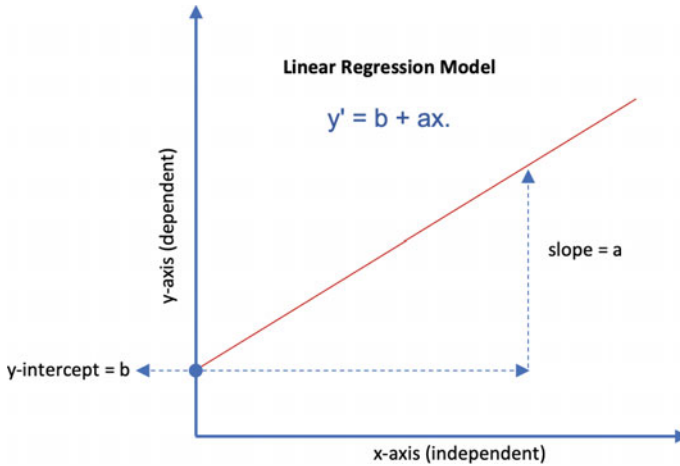
In theory, *variables* can be either dependent or independent based on how the researcher(s) design or sets up the experiments or hypothesis. It is important to note that the research design (whether qualitative or quantitative) resolves to or leads to what variables are manipulated (independent) or are consequently measured/affected (dependent) as a result of that manipulation. The authors will discuss this in detail in the following subsections.

#### 5.1.1.1 Independent Variable

*Independent variable*, otherwise referred to as the *manipulated variable(s)* represents the items or parameters being changed or manipulated by the researcher (Boyd, 2008; Shuttleworth & Wilson, 2008a, 2008b). According to Boyd (2008), an *independent variable* is expected to influence or at least be correlated with another variable (i.e., the dependent variable—see Sect. 5.1.1.2) in a data sample or scientific experiment. A lot of the time, the independent variable(s) are similarly referred to as the *controlled variable* as this is the variable(s) that are being decided (selected) or manipulated by the researchers in the experiment. In fact, an independent variable is considered “independent” because its distinctive attributes do not depend on the variation of other variables in a specific experiment or setup (Shuttleworth & Wilson, 2008a, 2008b). Hence, an “independent variable” is the variable that its value or change is not affected by other variables in an experiment.

Technically, as gathered in Fig. 5.1 (see more detail in Chap. 6); in linear regression models, *x-axis* (horizontal axis) is normally used to represent the *independent variable* in a graph or equation.

In principle, when we consider the graph (Fig. 5.1) which shows a linear relationship between  $x$  and  $y$ , the value of  $y$  is represented as a function of  $x$  “ $f(x)$ ”. Meaning that  $y$ , the dependent variable (see: Sect. 5.1.1.2) is dependent upon the value of  $x$ . Consequentially, the final outcome or result of the formula can be interpreted as:  $y$  depends on the  $x$  value (i.e., the independent variable) which can be changed or changes on its own. According to Sarikas (2020), a typical example of independent variables is studies or samples representing age and time. Basically, there is nothing the researcher or any likely factors could neither do to slow or speed up the time, nor decrease or increase the age. Thus, it can be said that the variables ( $x$ ), e.g., age and time, are often independent of every other variable in scientific experiments.



**Fig. 5.1** Graph representing the independent (x) and dependent (y) variables

### 5.1.1.2 Dependent Variables

As illustrated earlier in Fig. 5.1,  $y$  (i.e., vertical axis) is normally used to represent the *dependent variable* in a graph or equation. The authors highlighted in the figure (Fig. 5.1), how the independent variables ( $x$ ) relates to the dependent variable ( $y$ ). This is given that quite often either of the variables cannot be defined or discussed without referring to the other. Here, we explain the dependent variable in detail. Earlier on, we noted that in a typical experiment or hypothesis testing, the independent variables represent the items or parameters that are being manipulated, and, in turn, (the effects) subsequently observed or recorded. By definition, those observed or recorded effects are referred to as the *dependent variables* (Cao, 2008; Shuttleworth & Wilson, 2008a, 2008b). According to Shuttleworth and Wilson (2008a, 2008b), the *dependent variables* are often the *hypothesized consequences* (effects) as a result of manipulating the independent variable(s). Therefore, in a typical experiment, the dependent variable is assumed to respond to the independent variable. Thus, in theory, the dependent variables are also referred to as the “response variables”. Cao (2008) states that the judgment to treat a variable as a “dependent” might not only mean that an independent variable predicts the said variable but also happens to cause (or effects) the dependent variable.

### 5.1.1.3 Independent Versus Dependent Variables

In the following figure (Fig. 5.2), the authors show that the *independent variable* is the variable the researchers change (or controls), and it is expected that it will have a direct effect on the dependent variable. Thus, the *dependent variable* is the variable

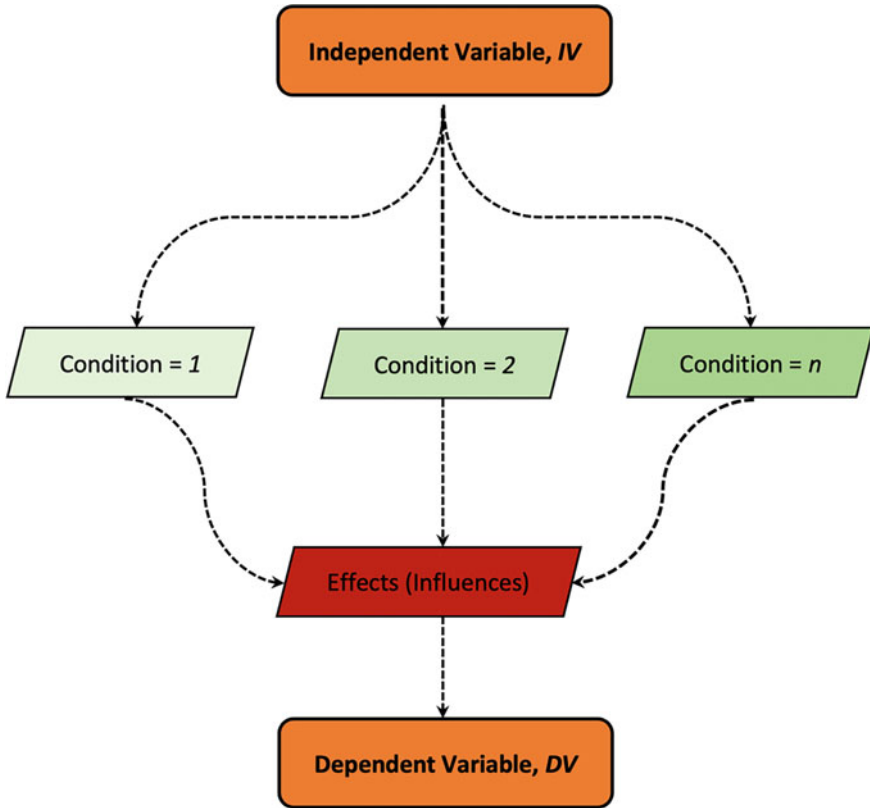


Fig. 5.2 Relationship and condition between independent versus dependent variables

being tested (or measured) during the experimentation, hence, as the name implies is “dependent” on the independent variable (McLeod, 2019).

The illustration in Fig. 5.2 shows that the purpose of any typical research experimentation or hypothesis testing should be focused on determining possible effects (influence) that leads to the *dependent variable* (DV) which may be caused by changing or altering (conditions) the *independent variables* (IV).

Furthermore, the authors provides in Table 5.1 some of the distinctive features of the independent (IV) versus the dependent variables (DV) that may guide the work of the researchers in determining the different types or categories of variables when conducting their experiments.

In summary, the independent variable (IV) provides the “input” to the statistical test or hypothesis which is modified by the model or the adopted method(s) by the researchers to change the “output” (dependent variable, DV). Interestingly, Cao (2008) notes that to conclude whether the dependent variable (DV) is caused by the independent variable (IV); that it is important to establish the relationship between

**Table 5.1** Independent versus dependent variable

| Independent (IV)                                       | Dependent (DV)   |
|--|--|
| Manipulated variable                                   | Response variable  |
| X or horizontal axis                                   | Y or vertical axis                                       |
| What the researchers change or what changes on its own | What is being studied or is being measured or determined |
| Input of an experiment                                 | Output of an experiment                                  |
| Causes the variation                                   | Depends on the variation                                 |

the two variables based on some pre-defined criteria. For example, as the authors outline below:

- The two variables (independent versus dependent) must be correlated, thus, a change in one variable (mainly the independent variable) must be accompanied by a change in the other (dependent variable).
- The observed correlation between the variables (independent versus dependent) must be genuine and conform to the measures by validity. In other words, the resultant relationship cannot be explained by other variables. Although the independent variable can be one of many other factors that could influence the dependent variable.
- Causal relationships between the independent and dependent variables are typically probabilistic in nature rather than deterministic; meaning that such association will not always necessarily be true for every run test or case scenario.
- When considering the order or timing of occurrence, the dependent variable (DV) must follow the independent variable (IV). For instance, a researcher who seeks to determine how ones' level of education influences ones' level of academic performance or work production would necessarily show that changes in the latter (academic or work production) occurred after changes in the former (education level).

### ***5.1.2 Examples and Use Case Scenarios of Independent Versus Dependent Variables***

In the preceding section (Sect. 5.1.1), we described the different types of variables (independent and dependent) including the distinctive features or how to identify each type in a dataset or experiments. Here, in Table 5.2, the authors look at some examples or use case scenarios that can guide the researchers or statisticians in defining or establishing the differences between the two variables (IV versus DV) especially in the different research settings or domains that they (researchers, statisticians, data analyst) undergo.

**Table 5.2** Use case scenarios for identifying independent versus dependent variables

| Research scenarios   | Independent variable (IV)   | Dependent variable (DV)   |
|--|---|---|
| Study that focuses on determining the impact of a particular teaching/learning method  | <ul style="list-style-type: none"> <li>• The teaching/learning method</li> </ul>                | <ul style="list-style-type: none"> <li>• Impact of the method</li> </ul>                                    |
| Study that explores how much resources is invested to acquire some certain books for a library   | <ul style="list-style-type: none"> <li>• How many books were acquired</li> </ul>                | <ul style="list-style-type: none"> <li>• Resources invested</li> </ul>                                      |
| An experiment that tests if the changes in the location of a certain work environment affect the outputs or productivity               | <ul style="list-style-type: none"> <li>• Change in locations</li> </ul>                         | <ul style="list-style-type: none"> <li>• Whether the productivity is affected or not</li> </ul>             |
| Sports experiment to determine if the number of hours spent training improves the athlete’s performance                                | <ul style="list-style-type: none"> <li>• Number of hours spent training</li> </ul>              | <ul style="list-style-type: none"> <li>• Athletes performance</li> </ul>                                    |
| A researcher sets out to determine which type of educational model mostly supports the students in meeting their learning goals        | <ul style="list-style-type: none"> <li>• The types of educational model</li> </ul>              | <ul style="list-style-type: none"> <li>• Level of support or meeting the students’ learning goal</li> </ul> |
| Medical experiment that focuses on determining how increase in the temperature of patients affects their response to treatment         | <ul style="list-style-type: none"> <li>• Patient’s temperature</li> </ul>                       | <ul style="list-style-type: none"> <li>• Response to treatment</li> </ul>                                   |
| Agricultural research that seeks to establish how different factors (e.g., sunlight, water, etc.) affects the growth of certain plants | <ul style="list-style-type: none"> <li>• The different factors, sunlight, water, etc</li> </ul> | <ul style="list-style-type: none"> <li>• How tall the plants grow</li> </ul>                                |
| A social science research that investigates how the level of ones’ education affects ones’ maturity or level of thought                | <ul style="list-style-type: none"> <li>• Level of education</li> </ul>                          | <ul style="list-style-type: none"> <li>• Level of thought</li> </ul>  |

## 5.2 Summary

The relationship between the *independent* (IV) and *dependent* (DV) variables is the key foundation of most statistical data analysis or scientific tests. An ample understanding or identification of the independent versus dependent variables is paramount to having a good knowledge about the outcome or impact of the scientific research or

how the experimentations are being done/conducted. In typical scientific research, the researchers can establish whether there is a significant correlation between the two variables (IV versus DV). In turn, the outcome of the procedures or methods (tests) enables the investigators to draw conclusions by either accepting or rejecting the pre-defined research hypothesis.

Quite often, it can be anything from really cumbersome or easy to identify the independent (IV) and dependent (DV) variables for any research study or data sample. According to Sarikas (2020), an easy way to identify the independent or dependent variable during an experiment is: independent variables (IV) are what the researchers change or changes on its own, whereas dependent variables (DV) are what changes as a result of the change in the independent variable (IV). Thus, independent variables (IV) are the *cause* while dependent variables (DV) are the *effect*.

Finally, it is important to remember that while some studies are likely to have one dependent variable (DV) and one independent variable (IV), it is also possible to have several of each type of the variables, i.e, more than one independent or dependent variables, in an experiment, as we will illustrate in Chapter 6 with some examples. Researchers might also want to learn or explore how variables in a single independent variable or factor affect several distinct dependent variables, or vice and versa (Cherry, 2019).

## References

- Agravante, M. (2018). *What Is the Meaning of Variables in Research?* <https://sciencing.com/measuring-variables-research-6164255.html>.
- Boyd, H. H. (2008). *Independent Variable In: Encyclopedia of Survey Research Methods*. <https://doi.org/10.4135/9781412963947>.
- Cao, X. (2008). *Dependent Variable In: Encyclopedia of Survey Research Methods*. <https://doi.org/10.4135/9781412963947>.
- Cherry, K. (2019). *Dependent Variable in Experiments*. <https://www.verywellmind.com/what-is-a-dependent-variable-2795099>.
- Dul, J. (2016). *Necessary Condition Analysis (NCA): Logic and Methodology of “Necessary but Not Sufficient” Causality*, 19(1), 10–52. <https://doi.org/10.1177/1094428115584005>.
- McLeod, S. (2019). *What are Independent and Dependent Variables?* Simply Psychology. <https://www.simplypsychology.org/variables.html>.
- Okoye, K., Tawil, A. R. H., Naeem, U., Bashroush, R., & Lamine, E. (2014). A semantic rule-based approach supported by process mining for personalised adaptive learning. *Procedia Computer Science*. <https://doi.org/10.1016/j.procs.2014.08.031>
- Sarikas, C. (2020). *Independent and Dependent Variables: Which Is Which?* <https://blog.prepscholar.com/independent-and-dependent-variables>.
- Shuttleworth, Martyn, & Wilson, L. T. (2008). *Dependent Variable*. <https://explorable.com/dependent-variable>
- Shuttleworth, M., & Wilson, L. T. (2008). *Independent Variable*. <https://explorable.com/independent-variable>.
- Stone-Romero, E. F. (2002). The relative validity and usefulness of various empirical research designs. In *Handbook of research methods in industrial and organizational psychology*. (pp. 77–98). Blackwell Publishing.



- Tynan, M. C., Credé, M., & Harms, P. D. (2020). Are individual characteristics and behaviors necessary-but-not-sufficient conditions for academic success?: A demonstration of Dul's (2016) necessary condition analysis. *Learning and Individual Differences*, 77 (December 2019), 101815. <https://doi.org/10.1016/j.lindif.2019.101815>.