

1.2: Theory and Empirical Research

This book is concerned with the connection between theoretical claims and empirical data. It is about using statistical modeling; in particular, the tool of regression analysis, which is used to develop and refine theories. We define **theory** broadly as a set of interrelated propositions that seek to explain and, in some cases, predict an observed phenomenon.

***Theory:** A set of interrelated propositions that seek to explain and predict an observed phenomenon.*

Theories contain three important characteristics that we discuss in detail below.

Characteristics of Good Theories

- *Coherent and internally consistent*
- *Causal in nature*
- *Generate testable hypotheses*

1.2.1 Coherent and Internally Consistent

The set of interrelated propositions that constitute a well-structured theory are based on **concepts**. In well-developed theories, the expected relationships among these concepts are both coherent and internally consistent. **Coherence** means the identification of concepts and the specified relationships among them are logical, ordered, and integrated. An **internally consistent** theory will explain relationships with respect to a set of common underlying causes and conditions, providing for consistency in expected relationships (and avoidance of contradictions). For systematic quantitative research, the relevant theoretical concepts are defined such that they can be measured and quantified. Some concepts are relatively easy to quantify, such as the number of votes cast for the winning Presidential candidate in a specified year or the frequency of arrests for gang-related crimes in a particular region and time period. Others are more difficult, such as the concepts of democratization, political ideology or presidential approval. Concepts that are more difficult to measure must be carefully **operationalized**, which is a process of relating a concept to an observation that can be measured using a defined procedure. For example, political ideology is often operationalized through public opinion surveys that ask respondents to place themselves on a Likert-type scale of ideological categories.

Concepts and Variables

A concept is a commonality across observed individual events or cases. It is a regularity that we find in a complex world. Concepts are our building blocks to understanding the world and to developing theory that explains the world. Once we have identified concepts we seek to explain them by developing theories based on them. Once we have explained a concept we need to define it. We do so in two steps. First, we give it a dictionary-like definition, called a nominal definition. Then, we develop an operational definition that identifies how we can measure and quantify it.

Once a concept has been quantified, it is employed in modeling as a **variable**. In statistical modeling, variables are thought of as either dependent or independent variables. A **dependent variable**, Y , is the outcome variable; this is the concept we are trying to explain and/or predict. The **independent variable(s)**, X , is the variable(s) that is used to predict or explain the dependent variable. The expected relationships between (and among) the variables are specified by the theory.

Measurement

When measuring concepts, the indicators that are used in building and testing theories should be both **valid** and **reliable**. Validity refers to how well the measurement captures the concept. Face validity, for example, refers to the plausibility and general acceptance of the measure, while the domain validity of the measure concerns the degree to which it captures all relevant aspects of the concept. Reliability, by contrast, refers to how consistent the measure is with repeated applications. A measure is reliable if, when applied to the repeated observations in similar settings, the outcomes are consistent.

Assessing the Quality of a Measure

Measurement is the process of assigning numbers to the phenomenon or concept that you are interested in. Measurement is straight-forward when we can directly observe the phenomenon. One agrees on a metric, such as inches or pounds, and then figures out how many of those units are present for the case in question. Measurement becomes more challenging when you cannot directly observe the concept of interest. In political science and public policy, some of the things we want to measure are directly

observable: how many dollars were spent on a project or how many votes the incumbent receives, but many of our concepts are not observable: is issue X on the public's agenda, how successful is a program, or how much do citizens trust the president. When the concept is not directly observable the operational definition is especially important. The operational definition explains exactly what the researcher will do to assign a number for each subject/case.

In reality, there is always some possibility that the number assigned does not reflect the true value for that case, i.e., there may be some error involved. Error can come about for any number of reasons, including mistakes in coding, the need for subjective judgments, or a measuring instrument that lacks precision. These kinds of error will generally produce inconsistent results; that is, they reduce reliability. We can assess the reliability of an indicator using one of two general approaches. One approach is a test-retest method where the same subjects are measured at two different points in time. If the measure is reliable the correlation between the two observations should be high. We can also assess reliability by using multiple indicators of the same concept and determining if there is a strong inter-correlation among them using statistical formulas such as Cronbach's alpha or Kuder-Richardson Formula 20 (KR-20).

We can also have error when our measure is not valid. Valid indicators measure the concept we think they are measuring. The indicator should both converge with the concept and discriminate between the concept and similar yet different concepts. Unfortunately, there is no failsafe way to determine whether an indicator is valid. There are, however, a few things you can do to gain confidence in the validity of the indicator. First, you can simply look at it from a logical perspective and ask if it seems like it is valid. Does it have face validity? Second, you can see if it correlates well with other indicators that are considered valid, and in ways that are consistent with theory. This is called construct validity. Third, you can determine if it works in the way expected, which is referred to as predictive validity. Finally, we have more confidence if other researchers using the same concept agree that the indicator is considered valid. This consensual validity at least ensures that different researchers are talking about the same thing.

Measurement of Different Kinds of Concepts

Measurement can be applied to different kinds of concepts, which causes measures of different concepts to vary. There are three primary **levels of measurement**; ordinal, interval, and nominal. **Ordinal level** measures indicate relative differences, such as more or less, but do not provide equal distances between intervals on the measurement scale. Therefore, ordinal measures cannot tell us *how much* more or less one observation is than another. Imagine a survey question asking respondents to identify their annual income. Respondents are given a choice of five different income levels: \$0-20,000, \$20,000-50,000, \$50,000-\$100,000, and \$100,000+. This measure gives us an idea of the rank order of respondents' income, but it is impossible for us to identify consistent differences between these responses. With an **interval level** measure, the variable is ordered and the differences between values are consistent. Sticking with the example of income, survey respondents are now asked to provide their annual income to the nearest ten thousand dollar mark (e.g., \$10,000, \$20,000, \$30,000, etc.). This measurement technique produces an interval level variable because we have both a rank ordering and equal spacing between values. Ratio scales are interval measures with the special characteristic that the value of zero (0) indicates the absence of some property. A value of zero (0) income in our example may indicate a person does not have a job. Another example of a ratio scale is the Kelvin temperature scale because zero (0) degrees Kelvin indicates the complete absence of heat. Finally, a **nominal level** measure identifies categorical differences among observations. Numerical values assigned to nominal variables have no inherent meaning, but only differentiate one type" (e.g., gender, race, religion) from another.

1.2.2 Theories and Causality

Theories should be causal in nature, meaning that an independent variable is thought to have a causal influence on the dependent variable. In other words, a change in the independent variable *causes* a change in the dependent variable. Causality can be thought of as the motor" that drives the model and provides the basis for explanation and (possibly) prediction.

The Basis of Causality in Theories

1. Time Ordering: The cause precedes the effect, $X \rightarrow Y$
2. Co-Variation: Changes in X are associated with changes in Y
3. Non-Spuriousness: There is not a variable Z that causes both X and Y

To establish causality we want to demonstrate that a change in the independent variable is a necessary and sufficient condition for a change in the dependent variable (though more complex, interdependent relationships can also be quantitatively modeled). We can think of the independent variable as a treatment, τ , and we speculate that τ causes a change in our dependent variable, Y. The gold

standard” for causal inference is an experiment where a) the level of τ is controlled by the researcher and b) subjects are randomly assigned to a treatment or control group. The group that receives the treatment has outcome Y_1 and the control group has outcome Y_0 ; the treatment effect can be defined as $\tau = Y_1 - Y_0$. Causality is inferred because the treatment was only given to one group, and since these groups were randomly assigned other influences should wash out. Thus the difference $\tau = Y_1 - Y_0$ can be attributed to the treatment.

Given the nature of social science and public policy theorizing, we often can’t control the treatment of interest. For example, our case study in this text concerns the effect of political ideology on views about the environment. For this type of relationship, we cannot randomly assign ideology in an experimental sense. Instead, we employ statistical controls to account for the possible influences of confounding factors, such as age and gender. Using multiple regression we control for other factors that might influence the dependent variable.¹

1.2.3 Generation of Testable Hypothesis

Theory building is accomplished through the testing of hypotheses derived from theory. In simple form, a theory implies (sets of) relationships among concepts. These concepts are then operationalized. Finally, models are developed to examine how the measures are related. Properly specified hypotheses can be tested with empirical data, which are derived from the application of valid and reliable measures to relevant observations. The testing and re-testing of hypotheses develops levels of confidence that we can have for the core propositions that constitute the theory. In short, empirically grounded theories must be able to posit clear hypotheses that are testable. In this text, we discuss hypotheses and test them using relevant models and data.

As noted above, this text uses the concepts of political ideology and views about the environment as a case study in order to generate and test hypotheses about the relationships between these variables. For example, based on popular media accounts, it is plausible to expect that political conservatives are less likely to be concerned about the environment than political moderates or liberals. Therefore, we can pose the **working hypothesis** that measures of political ideology will be systematically related to measures of concern for the environment – with conservatives showing less concern for the environment. In classical hypothesis testing, the working hypothesis is tested against a **null hypothesis**. A null hypothesis is an implicit hypothesis that posits the independent variable has no effect (i.e., null effect) on the dependent variable. In our example, the null hypothesis states ideology has no effect on environmental concern.

This page titled [1.2: Theory and Empirical Research](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Jenkins-Smith et al. \(University of Oklahoma Libraries\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.