2

Variables and Measurement

Variables and Data

Variables and Hypotheses

An important part of scientific research is forming a **hypothesis**—a testable statement about the relationship between two or more variables. A **variable** is a logical grouping of attributes that can be observed and measured and is expected to vary from person to person in a population. In social science research, variables have two important properties—they are exhaustive and mutually exclusive.

First, in order for a variable to be **exhaustive**, there must be a comprehensive list of the attributes that make up the variable. In order to measure marital status, for example, one would need a complete list of marital statuses that make up the variable. If a researcher looking to measure individuals' marital satisfaction asked individuals to report their marital status as either married or not married, this would not be an exhaustive representation of the variable. In particular, limiting the measurement to only these two attributes does not effectively measure the marital status of individuals who are divorced, separated, widowed, or have never been married. Thus, an exhaustive variable provides a more comprehensive account of marital status categories. Second, in order to be **mutually exclusive**, an individual score can be in only one response category and no others. In the original example, if a separated individual marked both "married" and "not married," then the response is not mutually exclusive. When a score contains more than one attribute of a variable, then a survey respondent might be counted twice for a given attribute.

This problem is discussed in the context of frequencies and distributions in Chapter 4.

Independent and Dependent Variables

An important part of hypothesis development is identifying or speculating about *how* the relationship is hypothesized to exist between two variables—which is the independent variable and which is the dependent variable. Generally, we assume that the **independent variable (IV)** influences, or leads to some change in, the **dependent variable (DV)**.[1] Thinking of it another way, a change in the dependent variable is *dependent* upon the independent variable. As discussed in Chapter 1, the scientific rationale for this designation is largely rooted in theory and logic. However, even if a researcher does not have a clear idea of which is the independent and which is the dependent variable, even the most basic bivariate statistical procedures (e.g., tables and graphs) require that a designation between the two exists. For the sake of clarity, in addition to new and important terms being presented in **bold** font, this chapter uses a separate font to distinguish `variables` from other text.

Drawing on theory and data about individuals' life satisfaction (`lifesatis`) and their self-reported number of close friends (`friends`), one might hypothesize that life satisfaction *depends* upon the number of close friends one has. In other words, we are proposing that variations in `friends`—the independent variable—will lead to changes in life satisfaction (`lifesatis`)—the dependent variable.

Number of Close Friends (`friends`) (IV) → Life Satisfaction (`lifesatis`) (DV)

However, depending on the topic of investigation, the opposite designations could also be appropriate. Perhaps individuals who are satisfied with their lives are more likely to socialize and make friends. In this case, the hypothesis would suggest that `lifesatis` is the *independent* variable and `friends` is the *dependent* variable.

Life Satisfaction (`lifesatis`) (IV) → Number of Close Friends (`friends`) (DV)

Application to Statistics and Statistical Interpretation: Many commonly used statistical procedures—and most of those discussed later in this book—will be framed around which variable is the independent variable and which

[1] When you are reading through and interpreting statistical analyses in published research articles, you might come across other commonly used terms for these concepts. Here are some of the other names used to describe independent and dependent variables:

Dependent Variable: Outcome Variable, Response Variable, Criterion Variable
Independent Variable: Predictor Variable, Explanatory Variable, Experimental Variable, Stimulus

is the dependent variable. This designation is also important for interpreting bivariate tables (Chapter 4) and data visualizations (Chapter 5).

Directional Relationships

When presenting a hypothesis, researchers also identify which, if any, direction they expect the relationship to operate. Therefore, in addition to explaining *how* the relationship is expected to exist (i.e., which variable is independent versus dependent), a researcher should also identify the *type* of relationship as either positive, negative, or nondirectional.

A hypothesized **positive relationship** between variables is one where both variables are expected to operate in the same direction (either up or down) together. In the previous example, we might hypothesize that individuals with more close `friends` will report higher life satisfaction. This is an example of a hypothesized positive relationship—the hypothesis suggests that a higher number of close friendships (`friends`) will be associated with higher scores on self-reported life satisfaction (`lifesatis`). This is essentially the same as saying that lower numbers of close friendships will be related to lower scores on life satisfaction. Since both variables are expected to operate in the same direction, the hypothesized relationship is positive. In the simplest terms, one might even say, "I hypothesize a positive relationship between `friends` and `lifesatis`."

A hypothesized **negative relationship** between two variables is one where both variables are expected to operate *in opposite directions*—as one increases, the other decreases—and vice versa. Depending on theory and logic, a researcher might also propose that maintaining an extensive close social network could be overwhelming and intensely stressful compared to someone with a small, close-knit group of friends. In this case, more friends could lead to less satisfaction—a negative relationship. A higher number of close friendships (`friends`) will be associated with lower life satisfaction (`lifesatis`). Accordingly, a lower number of close friendships will lead to higher life satisfaction.

Some hypotheses are **nondirectional**. If the study is exploratory and there is no reason to hypothesize a directional relationship, then the hypothesized relationship is left open. A researcher might hypothesize that there is a relationship between the size of someone's `social network` and the `number of dates` he or she goes on each week. However, she might choose to leave this hypothesis *nondirectional* if she is unsure of the direction of the relationship. For instance, those with many friends might be too busy in their social lives to go on many dates. On the other hand, those with a large social circle might have more opportunities to go on frequent dates with new people.

Some types of variables are not suited for directional hypotheses. The relationship between marital status (`marstat`) and number of close friends (`friends`) would be an example of this. Try and imagine the hypothesized direction between these two variables. It would be impossible because `marital` `status` does not operate on a continuum—it does not increase or decrease. In other words, it has a different *measurement*. This is one reason why understanding how variables are measured, or their "level of measurement," is an extremely important element in hypothesis development, research design, data analysis, and the interpretation of results.

Levels of Variable Measurement

One way of looking at the way variables are measured is based on the characteristics of the variables—whether, and how, they operate on a continuum. There are two primary levels of measurement for variables (with other subcategories within them based on additional criteria): categorical and quantitative.[2] These levels of measurement are key to guiding all statistical analyses because some types of variables have attributes that others do not.

Categorical Variables

Categorical variables are based on a series of categories that do not have meaningful numbers associated with them. There are several types of categorical variables, each discussed below: nominal variables, dichotomous variables, and ordinal variables.

Nominal Variables

Nominal variables are simply a list of different categories that cannot be rank ordered in any way. They are used to describe membership in mutually exclusive categories, but aside from assignment into a particular group, nominal variables have no other properties. `Marital status` (never married, married, divorced, separated, widowed) is an example of a commonly used nominal variable—individuals can be in one `marital status` category or another and nothing else is known about the categories. As the lowest level of measurement, nominal variables have the least amount of information; therefore, analysis and interpretation with nominal variables are limited to the most basic statistical analyses.

Given that quantitative social science data are coded numerically, the nominal variables are usually represented by arbitrary numbers (i.e., the number 1 could just as easily represent "never married" as the number 34 or 83 or 5). The number is simply a numerical identification code for that characteristic. The

---

[2] In your research, you might come across other terms that are used to identify the different levels of variable measurement:

Nominal: Categorical, String
Ordinal: Rank Order
Quantitative: Interval/Ratio, Continuous, Equal-Interval

meaning of those numbers is recorded in a **codebook**, a document that provides details on numerical codes and measurement to help researchers conduct and interpret their statistical analyses. For an example of a codebook entry, see Figure 2.1.

Dichotomous Variables

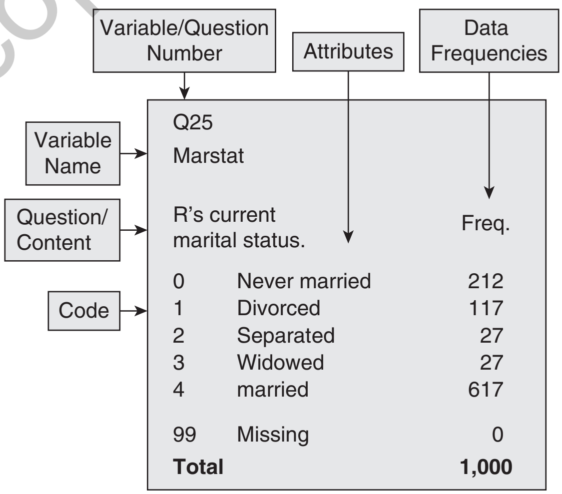
Dichotomous variables are those that have only two responses. While these variables might technically be considered nominal (since one either places in the category or not), they are often considered a special case since their binary nature is an attribute that other measures do not have. Among the most commonly used dichotomous variables in the social sciences is `sex` (male or female). Other examples of dichotomous variables are measurements that yield yes/no or true/false responses. As later chapters will show, there are special statistical procedures that can be used to conduct analysis using dichotomous variables.

Ordinal Variables

Ordinal variables are categories that specify a specific characteristic of an individual or individuals *but can be rank ordered*, thereby giving information about an individual's placement relative to others on the scale. These variables have an added property of rank-ordered attributes from lowest to highest—these numbers are usually classified in quantitative data as sequential numbers. For

FIGURE 2.1 ● Example of a Codebook Entry

FIGURE 2.1 identifies different parts of a codebook entry used to keep details about variables, their codes, and other characteristics.



example, self-reported life satisfaction (`lifesatis`) is measured with an ordinal level of measurement if individuals are categorized as (1) very dissatisfied through (7) very satisfied. Because there is a logical order to the categories, we know that individuals who are *very satisfied* are more satisfied than those who report being *very unsatisfied*. However, an ordered scale is still limited in the information it provides because we have no information on just *how much more* satisfied they are—that information is only present in quantitative variables.

Quantitative (Interval/Ratio) Variables

Quantitative variables are the highest level of measurement. These variables have meaningful numbers associated with them that refer to specific quantities. For example, one's actual `age` is a quantitative variable because the numbers are quantitatively meaningful (e.g., 21, 17, 91). Unlike ordinal variables, such as `social class`, quantitative variables allow us to know *exactly how different* individuals are relative to others. We know that someone who is 21 years old is 4 years older than someone who is 17—because the spacing of the intervals is equal. Therefore, quantitative variables give us the most information. For additional clarification on levels of variable measurement, see Box 2.1. One way to differentiate a quantitative variable from other variable types is to ask, "Would I be able to plot the variable's characteristics like coordinates on a graph?" If the answer is no, then it is not a quantitative variable (see Figure 2.2).

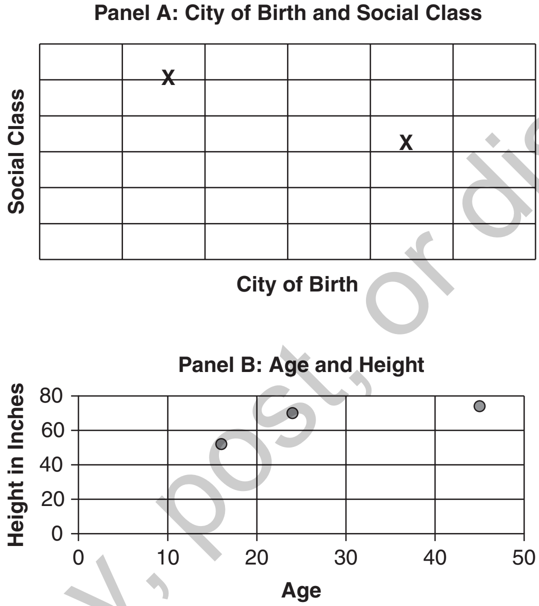
BOX 2.1

LEVELS OF VARIABLE MEASUREMENT AND PLAYING CARDS

A useful analogy for the difference between nominal, ordinal, and quantitative variables can be found in a standard 52-card deck of playing cards. First, in many card games, *the suits* (diamond, spade, club, and heart) are not rank ordered in any way. These categories are similar to nominal variables since they have no underlying assumptions about ordering or sequence—they are just different categories. *Face cards*, on the other hand, can be logically rank ordered from lowest to highest (Jack, Queen, King, and Ace). With face cards, however, there is no consistent interval that exists between them—just a logical rank ordering. The remainder of the deck can be thought of as discrete quantitative variables because they range from 2 to 10, with equal intervals assumed to exist between each number. Additionally, the card colors represent a dichotomous level of measurement since there are only two possibilities: black and red.

FIGURE 2.2 ● Nominal, Ordinal, and Interval/Ratio Data

FIGURE 2.2 helps illustrate the difference between nominal, ordinal, and interval/ratio data. Using a Cartesian plane, it would be impossible to meaningfully plot coordinates for `city of birth`, a nominal variable, and `social class`, an ordinal variable (Panel A). On the other hand, one *can* plot coordinates for `age` and `height` because both are interval/ratio variables (Panel B).



The Zero-Point

An **interval variable** is a quantitative variable with a zero-point that is arbitrary. In other words, a zero does not necessarily imply the absence of the construct. On the other hand, a **ratio variable** has a meaningful zero, whereby a zero indicates that there is a complete absence of that variable. For the purposes of the statistical procedures and interpretations discussed in this book, the difference between interval and ratio variables is less important than their quantitative nature; therefore, these variables will be referred to as either quantitative or interval/ratio variables.

Continuous and Discrete Quantitative Variables

There are some additional properties of variables that researchers sometimes take into consideration when designing a study and analyzing data. One such property is based on the proportions that exist between values. A **continuous quantitative variable** has an infinite number of possible values between two units. Using `weight` as an example, the interval between 179 and 180 pounds is theoretically infinite, as the decimals can extend infinitely beyond 179.7898. On the other hand, a **discrete quantitative variable** can take on only fixed values that are positive integers. For example, `number of close friends` is a discrete quantitative variable—the value cannot be infinitely reduced and can only be represented by whole numbers.

Transforming Variable Types

There is a hierarchy across the different variable types based on the amount of information they provide about a concept (e.g., interval/ratio variables have more information than ordinal variables, which have more information than nominal variables). Often, researchers are interested in transforming a variable into one with different properties. Quantitative variables can be **transformed** into categorical variables—that is, we can take the information from a quantitative variable and make it into a variable with the characteristics of those variables with less information. However, we cannot take information from a categorical variable and make it quantitative. For example, knowing someone's exact `height` would allow a researcher to classify him or her as short, average, or tall (an ordinal scale), but knowing if someone is short, average, or tall would not allow a researcher to extrapolate the person's actual `height`.

There are two commonly accepted exceptions to this rule. The first is **Likert-type items**, where an individual chooses from a range of possible responses that reflect his or her feelings, knowledge, or attitudes. For example, a common Likert-type item ranges from (1) *strongly disagree* to (7) *strongly agree*. These items are sometimes treated as quantitative variables in statistical analyses and interpretation. As such, researchers assume that the single-unit difference between a (1) and a (2) ranking is roughly the same as the single-unit difference between a (4) and a (5) ranking. Another exception to the transformation is the use of midpoints to identify a specific numerical amount when only a range is known. For example, a variable that asks for individuals' `annual income` with ranges (e.g., $50,000–$60,000) can be transformed into a quantitative variable by taking the midpoint of each range to transform the ordinal range into a meaningful quantity (e.g., $55,000).

Application to Statistics and Statistical Interpretation: Later chapters demonstrate the importance of levels of measurement of dependent and independent variables for statistical analysis and interpretation. For now, it is important to take note that levels of variable measurement are among the most important aspects of research design and largely depend on the way the data collection instrument is constructed.

Types of Relationships

Causal Relationships

One important consideration when developing a hypothesis is whether or not research is testing for a **causal relationship**, one where the independent variable *causes* a change in the dependent variable. In order to confirm that a causal relationship exists, researchers must establish three criteria. The criteria are discussed using the following information:

Hypothesis: Individuals with more close friends are more satisfied with their lives.

Independent Variable: Number of Close Friends (`friends`)

Dependent Variable: Life Satisfaction (`lifesatis`)

Direction: Positive

First, the variables must be **correlated**, meaning they operate together in some way—they are associated. If we see that `friends` and `lifesatis` both change in some way in relation to the other, then we know that the variables are correlated in some way.

"`Friends` and `lifesatis` are related—when one changes, so does the other one."

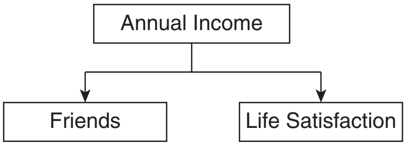
Second, **temporal precedence** must be established for the independent variable; that is, the independent variable (cause) must occur *before* the dependent variable (effect) in time. Thus, an increase in close friendships must precede an increase in life satisfaction ($X \rightarrow Y$).

"`Friends` increases *and then* `lifesatis` increases."

Third, and usually the most difficult criterion to establish, is that there should be no intervening factor that influences both variables. In this case, `income` might influence both variables, making it appear as though there is a causal relationship between `friends` and `lifesatis` where there is none (see Figure 2.3).

FIGURE 2.3 ● Causal Relationships

FIGURE 2.3 illustrates how income might be independently associated with both friends *and* happiness, making a relationship between the two *appear* causal.



"Individuals with higher incomes (`income`) have more close friends (`friends`)." "Individuals with higher incomes (`income`) have higher life satisfaction (`lifesatis`)."

Therefore, it might appear as though close friendships *cause* greater life satisfaction—but the relationship operates through income.

As discussed in Chapter 1, there are many different techniques to try to limit the effect of these intervening variables. In survey-based studies, researchers use longitudinal research designs in order to measure variables at two or more points in time. In such a study, researchers might explore changes in the size of friendship networks and life satisfaction over time.

When trying to establish a cause–effect relationship, experimental research is commonly considered the scientific gold standard. Under the proper experimental conditions—which includes randomization into control and experimental groups—intervening variables can be "controlled." Because selection into the experimental and control group was randomized, the control group is assumed to be like the experimental group in all other ways. Given that assumption, researchers can be more confident that they have eliminated the possibility of intervening factors, like `income` in the previous example.

Correlational Relationships

Some researchers are less interested in establishing a cause-and-effect relationship than simply exploring whether, how, and how strongly two variables are related. These are known as **correlational studies**, the most common of which is survey research. The primary goal of correlational research is to show whether a change in one variable is associated with change in another variable—and possibly identify the direction of the relationship.

Application to Statistics and Statistical Interpretation: Understanding whether or not the relationship between two variables is causal or correlational is important in order for researchers to avoid drawing erroneous or misleading conclusions by making causal statements when the relationship is not, in fact, a causal one. For example, if a survey research team were to interpret the results of its analysis and conclude, "Friends make you more satisfied with life," the findings would be misleading. If these researchers did not properly account (or "control") for other differences between individuals that can affect friendship and life satisfaction, such as income, then the relationship is only correlational, not causal.

Research Design and Measurement Quality

When dealing with variables, whether dependent or independent, measurement quality is an important element in the research process. Since social, behavioral, and health scientists often deal with concepts that are difficult to define and measure, there are several approaches to substantiate variables to improve their value for the research community.

Operationalization and Conceptualization

When researchers are testing hypotheses about dependent and independent variables, they must provide precise definitions of their main concepts and their measurement. Operationalization and conceptualization are the processes whereby researchers define and list the methods of observation for specific variables and measurements.

Conceptualization refers to the meaning, or conceptual definition, of a specific construct that a researcher proposes for her study. By providing a conceptual definition, the researcher is allowing readers to know exactly what she means when using that term. For example, if a study is examining whether `moving` leads to increases in children's `behavior problems`, one must first clearly define what is meant by "moving." Indeed, the term *moving* could mean any number of things from fine motor skills all the way to international migration. In order to conceptualize "moving," a researcher must provide a precise definition of what the term implies. In this case, `moving` means "a permanent relocation from one residence to another for longer than 1 year."

Conceptualization is the process of defining the concepts under study to clarify their meaning for the purposes of the research—and those meanings might differ from the conceptualization another researcher provides in another study. It would be confusing and counterproductive to conduct a study where no statement was made about what was meant by the terms being used. Perhaps others would recognize and understand `moving` to mean something completely different.

Operationalization is describing how the researcher empirically measures, or observes, the construct. While conceptualization entails providing a precise definition of a concept, the next step is to identify indicators of the measure. This process establishes the criteria being used to determine whether something exists versus when it does not. For example, a researcher might conceptualize a close friend as "a platonic associate who provides emotional support, instrumental support, or companionship." Operationalization is the process whereby a researcher describes the operations involved in observing or measuring the concept. Here, a researcher must describe *how* he will empirically observe whether a close friendship exists or not. Identifying a number of indicators, like those in Figure 2.4, would be one way to operationalize `close friendship`. In addition to the indicators in the figure, consider the multitude of alternate ways a researcher might operationalize close friendship.

Internal and External Validity

Researchers are always concerned about the legitimacy of their research design and results. As such, scientific research emphasizes the importance of

FIGURE 2.4 ● Conceptualizing and Operationalizing Close Friendship

FIGURE 2.4 presents one way of defining (i.e., conceptualizing) and making determinations about (i.e., operationalizing) close friendship. The following details are based on research on close friendships over the life cycle (Gillespie, Frederick, Harari, & Grov, 2015; Gillespie, Lever, Frederick, & Royce, 2015).

WHAT IS A CLOSE FRIEND?

A nonfamily platonic relationship where individuals provide emotional support, instrumental support, and companionship.

MEASURING CLOSE FRIENDSHIP

Expressive Support: You can talk with this person about intimate topics (i.e., sex life).
Instrumental Support: You can call on this person to help you if you are in trouble late at night.
Companionship: You expect this person to do something with you to celebrate your birthday.

internal and external validity. **Internal validity** is the degree to which a researcher can demonstrate that a causal relationship exists between variables. **External validity** refers to the applicability of a study to a wider, or more generalized, audience. Each of these concepts is discussed below, with some cautionary threats for each type.

Internal Validity

Internal validity is a concept used widely within experimental research to assert the legitimacy of a causal relationship between a dependent and an independent variable. As discussed above, researchers are often concerned with saying X *causes* Y; however, issues with the experiment can compromise researchers' ability to make such a causal statement. The following section highlights eight common problems that can occur in experiments that influence causal results. While some of these problems might seem unavoidable, researchers should try to reduce problems whenever possible.

What if a participant in an experiment knows the answers to posttest questions because he or she was asked the same questions in the pretest? This is known as a **testing effect.** For example, a statistics professor is interested in assessing how effective her `teaching methods` are by using a variation on the classical experimental design. She gives students a questionnaire (pretest) about `their math aptitude` at the beginning of the semester, teaches a `semester-long course on statistics`, and then gives the same exact test at the end of the semester (posttest). There is a possibility that students' responses on the posttest are higher not because of the professor's effective teaching methods, but because students remembered the questions, had time to think them over, or paid closer attention to areas they identified as difficult. In this case, the `professor's teaching` (independent variable/stimulus) would appear to be effective since the students' `math aptitude scores` were higher—but some of the difference could be related to the pretest matching the posttest.

What if I change the pretest and posttest to avoid this testing effect? If the same professor chooses to use a different posttest measurement at the end of the semester, not the same test as the pretest, there is a possibility of **instrumentation bias.** It is possible that the questions on the first test were easier, harder, or perceived as easier or harder than the earlier questions. In this sense, it would appear that students scored differently on `math aptitude` not because of the `information presented in the course`, but because of changes in the level of difficulty in the pretest and posttest instruments.

What if something important happens before the test is over? A research professor at Market University is interested in exploring how `participation in university groups` influences the size of students' `social networks`. He hypothesizes that the more students `participate in university clubs`, the larger their `social network` will be (i.e., a positive relationship where the independent variable is `participating in clubs` and the dependent variable is `size of social network`). Students in the experimental group must sign up for three social clubs on campus—the control group is advised to avoid joining groups for the duration of the experiment. However, if a flu outbreak influences the way students interact with one another, this epidemic could influence the outcome. This **history effect** influences the results of the experiment because a large-scale event (`flu outbreak`), not the stimulus (`participation in clubs`), influenced the outcome (`social network size`).

What if the people in the experimental and control groups are different from each other? This would be an example of **selectivity.** In the previous example, if the most gregarious students in the class are in the experimental group and those in the control group are the shyer and more reserved students, it might seem that there is a relationship between group affiliations and `social network size`. However, the relationship might be merely a reflection of this limitation in the experimental design. In research, this is known as

selection bias. This can happen when the experimental and control groups differ along some important characteristic related to the study.

What happens when people mature between the time the experiment begins and the time it ends? This is known as the **maturation effect.** Individuals might respond differently to a pretest than they otherwise would had time not passed. For example, individuals in the `group affiliation` and `social networks` experiment might have more friends at the end of the experiment simply because they have matured—perhaps over the course of the semester they developed different opinions on the importance of friendships and participation in university clubs. Again, this threat to internal validity creates the appearance of a causal relationship when one might not be present.

What happens when people leave the experiment? Experimental mortality occurs when people leave the experiment. Individuals can leave an experiment for a number of reasons (e.g., death, boredom, or moral disagreement with the subject matter). At best, this can limit the sample size; at worst, it can lead to selection bias. For example, if students in the experimental group to study the effects of `group affiliation` chose to leave because they are anxious about making new friends, then the experimental group might tend toward those more gregarious people who make friends easily.

What happens when the experimenter is biased because he or she knows who is in the experimental and who is in the control group? Experimenter bias occurs when the person running the experiment already knows who is in the experimental and control groups and then, perhaps unconsciously, treats them differently. In this case, the experimenter influences the interactions and feelings of the individuals in one group more so than the other. Researchers have developed a way to deal with the potential of this—they do what is called a **double-blind experiment,** which means that the experimenter does not know who is in the experimental or control groups.

What happens when people score very low or very high on the pretest? If the results vary because pretest scores were in the extreme regions, this might reflect **statistical regression toward the mean.** When individuals have very low scores at the pretest, there is a tendency to tend toward an average in the posttest. For example, students who score far below average on a pretest for `math aptitude` will tend to move closer toward the mean on the posttest. This makes it appear as though those students performed better because of the `professor's teaching effectiveness`, but they did so because students who scored in the lower extreme trended upward on the posttest. This threat to validity is based on the reliability of the instrument, or the ability to yield the same results with the same instrument after repeated measurement (discussed later).

Each of these threats to internal validity influences the way an experimenter draws conclusions about causal relationships—perhaps making it appear that the independent variable causes a change in the dependent variable when it does not. One way to improve internal validity is to create sound arguments and use caution when developing a rigorous research design. Another type of validity, external validity, points to whether the results of a given study are applicable to other groups and contexts.

External Validity

More often than not, social science researchers are interested in conducting studies and applying the results to larger groups. External validity is the extent to which the results of a study are applicable to other contexts. For example, a researcher who surveys his course to explore how `sleep habits` influence `school performance` would have low generalizability (external validity) since the sample consists only of college students. Therefore, the study's results might not apply to students at other grade levels—indeed, they might not even apply to students at other universities. Two ways to improve external validity is to (a) employ appropriate sampling procedures and (b) replicate the study on other groups to confirm the results. Additionally, researchers should aim to achieve high response rates and low dropout rates, which can help avoid a biased sample population.

Measurement Validity and Reliability

The previous section outlined two types of validity as they relate to the design of a study—whether researchers can confidently establish a causal relationship (internal validity) and whether the results apply to a wider group (external validity). However, in addition to the overall study design, researchers are also concerned with the quality of the measurements used to measure specific concepts. Measurement validity and reliability are ways that researchers inspect the quality of a measurement. Each, in its own respect, helps researchers assess whether their measurements are sound ones. However, keep in mind that these are not all-or-nothing assessments of a measurement. Instead, validity and reliability operate on a continuum from low to high—and researchers should strive to have both high validity and high reliability.

Measurement Validity

Measurement validity is an assessment of the quality of a measure to accurately tap into a target concept. There are four types of validity commonly encountered by researchers looking to assess the quality of their measures: face validity, content validity, criterion-based validity, and construct validity.

Does the measurement seem to be an accurate measurement of the concept? When a measurement seems to intuitively measure a target construct, the measure is presumed to have **face validity.** Face validity is a superficial assessment of whether or not the measurement "looks good at face value." Under some circumstances, low face validity might increase the overall measurement validity by helping avoid interviewer bias and social desirability.

Are all components of the construct being measured? If so, then the measurement has **content validity.** In order to be high in content validity, a measurement must assess the concept under study in a comprehensive way. This is especially important when measuring complex or nebulous concepts, such as `well-being`, where multiple dimensions are explored (e.g., physical wellness, emotional stability, life satisfaction) in order to form a measurement that comprehensively represents the entire construct.

Does the measurement correlate with other measures or outcomes? If so, it is thought to be high in **criterion-based validity.** Criterion-based validity is based on holding the measure against other criteria. When a measurement is determined to correlate with other theoretically relevant measures in the study, it has **concurrent validity.** If the measure corresponds to some theoretically relevant pre-established criterion (e.g., GRE scores and `graduate school performance`), then it has established **predictive validity.**

Is the instrument truly measuring the construct under study and not some other construct? If the measurement is an adequate assessment of the construct, then it is thought to have **construct validity.** For example, if researchers use an instrument to measure `self-esteem` in adolescence, they would need to confirm that their measure was tapping into self-esteem and not some other (possibly related) concept, such as depression, anxiety, or loneliness.

Researchers must be transparent by (a) demonstrating the effectiveness of a given measure and (b) identifying limitations and discussing the ways such limitations could influence the results.

Reliability

In addition to being accurate, measurements must also be consistent (i.e., have **reliability**). When a measure produces consistent results after repeated administration, it is reliable. For example, in survey research, if an open-ended response option is used to measure the `number of dates` a college student went on in a year's time, the answer could yield less reliable results given that a precise answer might not be known and a respondent might guess (which means it is not a very reliable measure). On the other hand, a set of ranges might help an individual provide an approximate response, leading to more consistency in the results. There are several ways to help ensure a measure is reliable. First, make sure questions are written clearly. Also, make sure closed-ended questions have a realistic number of response options from which to choose, and possibly an option for "don't know/NA."

Testing for Reliability

This section discusses several tests to assess the reliability of a measurement. First, **test–retest reliability** estimates a measurement's reliability based on the consistency of results after repeated administration. Responses from measures taken at Time 1 and Time 2 are assessed to estimate the stability of the measure over the two time points. A reliable question should elicit a similar response from one administration to the next. Second, **alternate forms reliability** refers to consistency between two different versions of a measure that probes the same construct. Third, **split-test reliability** groups similar items in a measurement instrument into two sets of equivalent items that are split into two halves. The scores from each half are compared to determine the degree of correlation between them. Correlation should be high among questions reliably measuring the same concept. Last, **Cronbach's alpha** is a statistical summary measure of the **internal consistency** of data collected across multiple items that form a scale. While there are no hard and fast rules regarding interpretation of Cronbach's alpha, the higher the value of Cronbach's alpha, the more consistent the items. For example, a Cronbach's alpha of .83 indicates that 83% of the variation is shared across the items.

Conclusion

In order to test a bivariate or multivariate hypothesis, one needs to define variables as either independent or dependent. If applicable, the direction of the hypothesis should also be stated, keeping in mind that certain types of variables (e.g., nominal) are not designed for directional relationships. Once a clear and testable hypothesis is developed, a research design must be chosen, keeping in mind both internal and external validity. Importantly, avoid hypothesizing a causal relationship if the research design is ill equipped to establish such relationships. After the research design has been chosen, a researcher must develop accurate and consistent measurements for the concepts under study. Each of these guidelines helps researchers target and address issues of precision that quality research must take into consideration. In order to fully grasp the concepts in the chapters to follow, an understanding of these concepts is necessary.

Terms

alternate forms reliability 25
categorical variables 12
causal relationship 17
codebook 13
conceptualization 19
concurrent validity 24
construct validity 24
content validity 24
continuous quantitative variable 16
correlated 17
correlational studies 18
criterion-based validity 24
Cronbach's alpha 25
dependent variable (DV) 10
dichotomous variables 13
discrete quantitative variable 16
double-blind experiment 22
exhaustive 9
experimental mortality 22
experimenter bias 22
external validity 20
face validity 24
history effect 21
hypothesis 9
independent variable (IV) 10
instrumentation bias 21
internal consistency 25
internal validity 20
interval variable 15
Likert-type items 16
maturation effect 22
measurement validity 23
mutually exclusive 9
negative relationship 11
nominal variables 12
nondirectional 11
operationalization 19
ordinal variables 13
positive relationship 11
predictive validity 24
quantitative variables 14
ratio variable 15
reliability 25
selection bias 22
selectivity 21
split-test reliability 25
statistical regression toward the mean 22
temporal precedence 17
testing effect 20
test–retest reliability 25
transformed 16
variable 9

References

Gillespie, B. J., Frederick, D., Harari, L., & Grov, C. (2015). Homophily, close friendship, and life satisfaction among gay, lesbian, heterosexual, and bisexual men and women. *PLoS One, 10*(6), e0128900.

Gillespie, B. J., Lever, J., Frederick, D., & Royce, T. (2015). Close adult friendships, gender, and the life cycle. *Journal of Social and Personal Relationships, 32*(6), 709–736.