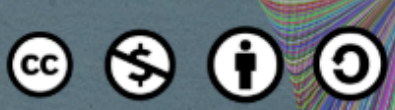
Quantitative Research Methods

for the Applied Human Sciences

Peter Morden



Quantitative Research Methods for the Applied Human Sciences Copyright © 2024 by Peter Morden is licensed under a ***Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License****, except where otherwise noted.*

Dr. Peter Morden customized the anonymized open textbook ***Principles of Sociological Inquiry: Qualitative and Quantitative Methods,*** originally licensed under a **CC-BY-NC-SA 3.0,** and adapted several chapters from additional open resources. Where relevant, copyright notices can be found at the chapter level.

This book was produced with Pressbooks (https://pressbooks.com) and rendered with Prince.

Contents

Accessibility Statement ix

CHAPTER 1: INTRODUCTION

How Do We Know What We Know? 2

Different Sources of Knowledge 2

Ontology and Epistemology 5

Science, Social Sciences, and Applied Human Sciences 6

The Science of Social Sciences 6

Conducting Ethical Research 9

Specific Ethical Issues to Consider 10

The Practice of Science Versus the Uses of Science 14

Doing Science the Ethical Way 14

Using Science the Ethical Way 15

Key Takeaways, Exercises and References 17

Key Takeaways 17

Exercises 17

References 19

CHAPTER 2: LINKING METHODS WITH THEORY

Micro, Meso, and Macro Approaches 22

Social Science at Three Different Levels 22

Paradigms, Theories, and How They Shape a Researcher's Approach 25

Paradigms in Social Science 25

Social Science Theories 27

Inductive or Deductive? Two Different Approaches 29

Revisiting an Earlier Question 33

Key Takeaways, Exercises and References 34

Key Takeaways 34

Exercises 34

References 35

CHAPTER 3: DEVELOPING A RESEARCH PROJECT

Starting Where You Already Are 40

How Do You Feel About Where You Already Are? 41

What Do You Know About Where You Already Are? 42

Is It Empirical? 42

Complementary Disciplines in Applied Human Sciences 44

Is It a Question? 45

Some Specific Examples 46

Feasibility 49

How to Design a Research Project? 50

Goals of the Research Project 50

Exploration, Description, Explanation 51

Idiographic or Nomothetic? 52

Applied or Basic? 52

Causality 54

Units of Analysis and Units of Observations 56

Errors in Logical Reasoning 58

Hypothesis 59

Key Takeaways, Exercises, and References 61

Key Takeaways 61

Exercises 62

References 63

CHAPTER 4: CONDUCTING A LITERATURE REVIEW

What is a Literature Review? 65

Literature Review Basics 65

Common Literature Review Errors 67

Synthesizing Literature: Putting the Pieces Together 68

Creating a topical outline 70

Additional resources for synthesizing literature 71

Writing the Literature Review 73

The problem statement 74

The structure of a literature review 74

Editing Your Literature Review 80

References 82

CHAPTER 5: CONCEPTUALIZATION AND OPERATIONALIZATION

Measurement 84

What Do Social Scientists Measure? 84

How Do Social Scientists Measure? 86

Conceptualization 87

Concepts and Conceptualization 87

A Word of Caution About Conceptualization 89

Operationalization 91

Indicators 91

Putting It All Together 93

Measurement Quality 94

Reliability 94

Validity 96

Complexities in Measurement 98

Levels of Measurement 98

Key Takeaways, Exercises, and References 100

Key Takeaways 100

Exercises 100

References 101

CHAPTER 6: SAMPLING

Population Versus Samples 104

Sampling Without Generalizing 105

Nonprobability Sampling 105

Types of Nonprobability Samples 106

Sampling for Generalizability 109

Probability Sampling 109

Types of Probability Samples 110

A Word of Caution: Questions to Ask About Samples 115

Who Sampled, How Sampled, and for What Purpose? 115

Key Takeaways, Exercises, and References 118

Key Takeaways 118

Exercises 118

References 119

CHAPTER 7: SURVEY RESEARCH

Survey Research: What Is It and When Should It Be Used? 122

Pros and Cons of Survey Research 123

Strengths of Survey Method 123

Types of Surveys 125

The Time Dimension 125

Administration 128

Designing Effective Questions and Questionnaires 131

Asking Effective Questions 131

Response Options 135

Designing Questionnaires 137

Response Rate 139

Key Takeaways, Exercises, and References 140

Key Takeaways 140

Exercises 140

References 141

CHAPTER 8: EXPERIMENTAL DESIGN

Experimental Design: What Is It and When Should It Be Used? 144

Pre-experimental and Quasi-Experimental Design 148

The Logic of Experimental Design 151

Threats to Validity 155

Key Takeaways, Exercises, and References 157

Key Takeaways 157

Exercises 158

References 158

CHAPTER 9: DESCRIPTIVE STATISTICS

Describing Single Variables 161

The Distribution of a Variable 161

Measures of Central Tendency and Variability 166

Central Tendency 166

Measures of Variability 167

Percentile Ranks and z Scores 168

Describing Statistical Relationships 170

Differences Between Groups or Conditions 170

Correlations Between Quantitative Variables 172

Expressing Your Results 174

Presenting Descriptive Statistics in Writing 174

Presenting Descriptive Statistics in Figures 175

Expressing Descriptive Statistics in Tables 178

Conducting Your Analyses 180

Prepare Your Data for Analysis 180

Preliminary Analyses 182

Understand Your Descriptive Statistics 183

Key Takeaways, Exercises, and References 184

KEY TAKEAWAYS 184

Exercises 185

References 186

CHAPTER 10: INFERENTIAL STATISTICS

Understanding Null Hypothesis Testing 189

The Logic of Null Hypothesis Testing 190

Role of Sample Size and Relationship Strength 192

Statistical Significance versus Practical Significance 193

Some Basic Null Hypothesis Tests 195

The t-Test 195

The Analysis of Variance 199

Testing Correlation Coefficients 201

Additional Considerations 204

Errors in Null Hypothesis Testing 204

Statistical Power 205

Criticisms of Null Hypothesis Testing 207

What to do? 209

Key Takeaways, Exercises, and References 210

Key Takeaways 210

Exercises 211

References 211

Accessibility Statement

This open textbook was designed to meet accessibility guidelines. Using available tools, we have checked to ensure it is compatible with screen readers in the online interactive Pressbooks and ebook versions. The guidebook is also available for download in EPUB, Digital PDF, and MOBI formats.

For the date of the last revision, see **Versioning History** in the back matter.

For more support or help with accessibility or any other aspect of this customized open textbook please contact:

CHAPTER 1: INTRODUCTION

Do you like to know things? Do you ever wonder what other people know or how they know what they do? Have you ever made a decision, and do you plan to make decisions in the future? If you answered yes to any of these questions, then you will probably find the information in this book—particularly the information on research methods—very useful. If you answered no to all of them, I suspect that you will have reconsidered by the time you finish reading this text. Let's begin by focusing on the information in this chapter. Here we'll consider the variety of ways that we know things and what makes social scientific knowledge unique. We'll also consider why any of this might matter to you and preview what's to come in later chapters.

LEARNING OBJECTIVES

- Define research methods.
- Identify and describe the various ways of knowing presented in this section.
- Understand the weaknesses of non-systematic ways of knowing.
- Define ontology and epistemology and explain the difference between the two.
- Define science.
- Describe the specific considerations of which social scientists should be aware.
- Be able to describe and discuss some of the reasons why students should care about social scientific research methods.
- Define the term human subjects.
- Define informed consent, and describe how it works.
- Identify the unique concerns related to the study of vulnerable populations.
- Understand the definitions of and the differences between anonymity and confidentiality.
- Describe what it means to use science in an ethical way.

How Do We Know What We Know?

If I told you that the world is flat, I'm hoping you would know that I'm wrong. But how do you know that I'm wrong? And why did people once believe that they knew that the world was flat? Presumably the shape of the earth did not change dramatically in the time that we went from "knowing" one thing about it to knowing the other; however, something certainly changed our minds. Understanding both what changed our minds (science) and how might tell us a lot about what we know, what we think we know, and what we think we can know.

This book is dedicated to understanding exactly how it is that we know what we know. More specifically, we will examine the ways that social scientists come to know social facts. Our focus will be on one particular way of knowing: social scientific research methods. ***Research methods*** are a systematic process of inquiry applied to learn something about our social world. But before we take a closer look at research methods, let's consider some of our other sources of knowledge.

Different Sources of Knowledge

What do you know about only children? Culturally, our stereotype of children without siblings is that they grow up to be rather spoiled and unpleasant. We might think that the social skills of only children will not be as well developed as those of people who were reared with siblings. However, sociological research shows that children who grow up without siblings are no worse off than their counterparts with siblings when it comes to developing good social skills (Bobbitt-Zeher & Downey, 2010). Social scientists consider precisely these types of assumptions that we take for granted when applying research methods in their investigations. Sometimes we find that our assumptions are correct. Often as in this case, we learn that the thing that everyone seems to know to be true isn't so true after all.

Many people seem to know things without having a scientific background relevant to the topic. Of course, there are a variety of ways we know things that don't involve scientific research methods. Some people know things through experiences they've had, but they may not think about those experiences systematically; others believe they know things based on selective observation or overgeneralization; still others may assume that what they've always known to be true is true simply because they've always known it to be true. Let's consider some of these alternative ways of knowing before focusing on social-scientific ways of knowing.

Many of us know things simply because we've experienced them directly. For example, you would know that electric fences can be pretty dangerous and painful if you touched one while standing in a puddle of water. We all probably have times we can recall when we learned something

because we experienced it. If you grew up in Quebec, you would observe plenty of kids learn each winter that it really is true that one's tongue will stick to metal if it's very cold outside. Similarly, if you passed a police officer on a two-lane highway while driving 50% over the speed limit, you would probably learn that that's a good way to earn a traffic ticket. So direct experience may get us accurate information but only if we're lucky (or unlucky, as in the examples provided here). In each of these instances, the observation process isn't really deliberate or formal. Instead, you would come to know what you believe to be true through informal observation. The problem with informal observation is that sometimes it is right, and sometimes it is wrong. And without any systematic process for observing or assessing the accuracy of our observations, we can never really be sure that our informal observations are accurate.

Suppose a friend of yours declared that "all men lie all the time" shortly after she'd learned that her boyfriend had told her a fib. The fact that one man happened to lie to her in one instance came to represent all experiences with all men. But do all men really lie all the time? Probably not. If you prompted your friend to think more broadly about her experiences with men, she would probably acknowledge that she knew many men who, to her knowledge, had never lied to her and that even her boyfriend didn't generally make a habit of lying. This friend committed what social scientists refer to as selective observation by noticing only the pattern that she wanted to find at the time. If, on the other hand, your friend's experience with her boyfriend had been her only experience with any man, then she would have been committing what social scientists refer to as overgeneralization, assuming that broad patterns exist based on very limited observations.

Another way that people claim to know what they know is by looking to what they've always known to be true. There's an urban legend about a woman who for years used to cut both ends off of a ham before putting it in the oven (Mikkelson & Mikkelson, 2005). She baked ham that way because that's the way her mother did it, so clearly that was the way it was supposed to be done. Her mother was the authority, after all. After years of tossing cuts of perfectly good ham into the trash, however, she learned that the only reason her mother ever cut the ends off ham before cooking it was that she didn't have a pan large enough to accommodate the ham without trimming it.

Without questioning what we think we know to be true, we may wind up believing things that are actually false. This is most likely to occur when an authority tells us that something is so (Adler & Clark, 2011). Our mothers aren't the only possible authorities we might rely on as sources of knowledge. Other common authorities we might rely on in this way are the government, our schools and teachers, and our churches and ministers.

Although it is understandable that someone might believe something to be true because someone he or she looks up to or respects has said it is so, this way of knowing differs from the sociological way of knowing, which is our focus in this text.

In sum, there are many ways that people come to know what they know. These include informal observation, selective observation, overgeneralization, authority, and research methods. Table 1.1 "Several Different Ways of Knowing" summarizes each of the ways of knowing described here. Of course, some of these ways of knowing are more reliable than others. Being aware of our sources of knowledge helps us evaluate the trustworthiness of specific bits of knowledge we may hold.

Table 1.1 Several Different Ways of Knowing

Way of Knowing	**Description**
Informal observation	Occurs when we make observations without any systematic process for observing or assessing accuracy of what we observed.
Selective observation	Occurs when we see only those patterns that we want to see or when we assume that only the patterns we have experienced directly exist.
Overgeneralization	Occurs when we assume that broad patterns exist even when out observations have been limited.
Authority	A socially defined source of knowledge that might shape our beliefs about what is true and what is not true.
Research Methods	An organized, logical way of learning and knowing about our social world.

Ontology and Epistemology

Thinking about what you know and how you know what you know involves questions of ontology and epistemology. Perhaps you've heard these terms before in a philosophy class; however, they are relevant to the work of social scientists, as well. As we begin to think about finding something out about our social world, we are probably starting from some understanding of what "is," what can be known about what is, and what the best mechanism happens to be for learning about what is.

Ontology deals with the first part of these sorts of questions. It refers to one's analytic philosophy of the nature of reality. In the social sciences, a researcher's ontological position might shape the sorts of research questions he or she asks and how those questions are posed. Some take the position that reality is in the eye of the beholder and that our job is to understand others' view of reality. Others feel that, while people may differ in their perception of reality, there is only one true reality. These scientists are likely to aim to discover that true reality in their research rather than discovering a variety of realities.

Like ontology, epistemology has to do with knowledge. But rather than dealing with questions about what is, epistemology deals with questions of how we know what is. In the social sciences, there are a number of ways to uncover knowledge. We might interview people to understand public opinion about some topic, or perhaps we'll observe them in their natural environment. We could avoid face-to-face interaction altogether by mailing people surveys for them to complete on their own or by reading what people have to say about their opinions in newspaper editorials. All these are ways that social scientists gain knowledge. Each method of data collection comes with its own set of epistemological assumptions about how to find things out. We'll talk in more depth about these ways of knowing, specifically ways that yield quantitative data for analysis, in later chapters.

Science, Social Sciences, and Applied Human Sciences

In "How Do We Know What We Know?" we considered a variety of ways of knowing and the philosophy of knowing. But this is not a philosophy text. In this section, we'll take a closer look at the science of social science and some specific considerations of which social scientists must be aware.

The Science of Social Sciences

The sources of knowledge we discussed in Section 1.1 "How Do We Know What We Know?" could have been labeled sources of belief. In the social sciences, however, our aim is to discover knowledge. Although we may examine beliefs in order to understand what they are and where they come from, ultimately we aim to contribute to and enhance knowledge. Science is a particular way of knowing that attempts to systematically collect and categorize facts or truths. A key word here is systematically; conducting science is a deliberate process. Unlike the ways of knowing described earlier, scientists gather information about facts in a way that is organized and intentional and usually follows a set of predetermined steps.

More specifically, researchers in social science disciplines use organized and intentional procedures to uncover facts or truths about people and society. The focus of social scientific research may be as small as individuals, couples or families, or as large as whole nations. The main point, however, is that social scientists study human beings in relation to one another and to their environments. In our next chapter, we'll explore how variations within the social sciences, such as one's theoretical perspective, may shape a researcher's approach. For now the important thing to remember is out focus on human social behavior and the scientific approach toward understanding that behavior.

Specific Considerations for the Social Sciences

One of the first and most important things to keep in mind is that social scientists aim to explain ***patterns in society***. Most of the time, a pattern will not explain every single person's experience, a fact that is both fascinating and frustrating. It is fascinating because, even though the individuals who create a pattern may not be the same over time and may not even know one another, collectively they create a pattern. Those new to the study of social behaviour may find these patterns frustrating because they may believe that the patterns that describe their gender, their age, or some other facet of their lives don't really represent their experience. Make no mistake: a pattern can exist among your cohort without your individual participation in it.

Let's consider some specific examples. One area commonly investigated, especially among social scientists and social-psychologists, is the impact of a person's social class background on his or her experiences in life. You probably wouldn't be surprised to learn that a person's social class background has an impact on his or her educational attainment and achievement. In fact, one group of researchers (Ellwood & Kane, 2000) in the early 1990s found that the percentage of children who did not receive any postsecondary schooling was four times greater among those in the lowest quartile income bracket than those in the upper quartile of income earners (i.e., children from high-income families were far more likely than low-income children to go on to university). Another recent study found that having more liquid wealth that can be easily converted into cash actually seems to predict children's math and reading achievement (Elliott, Jung, Kim, & Chowa, 2010).

These findings, that wealth and income shape a child's educational experiences, are probably not that shocking to any of us, even if we know someone who may be an exception to the rule. Sometimes the patterns that social scientists observe fit our commonly held beliefs about the way the world works. When this happens, we don't tend to take issue with the fact that patterns don't necessarily represent all people's experiences. But what happens when the patterns disrupt our assumptions?

For example, did you know that some research has shown that teachers are far more likely to encourage boys to think critically in school by asking them to expand on answers they give in class and by commenting on boys' remarks and observations? When girls speak up in class, teachers are more likely to simply nod and move on. The pattern of teachers engaging in more complex interactions with boys means that boys and girls do not receive the same educational experience in school (Sadker & Sadker, 1994). You and your classmates, both men and women, may find this news upsetting.

Objectors to these findings tend to cite evidence from their own personal experience, refuting that the pattern actually exists. The problem with this response, however, is that objecting to a social pattern on the grounds that it doesn't match one's individual experience misses the point about patterns. The point isn't that there are exceptions; the point is that there is a general rule that seems operative in different contexts and with different people.

Another matter that social scientists must consider is where they stand on the value of basic as opposed to applied research. In essence, this has to do with questions of for whom and for what purpose research is conducted. We can think of basic and applied research as resting on either end of a continuum.***Basic research*** is research for the sake of the discipline. Nothing more, nothing less. Sometimes researchers are motivated to conduct research simply because they happen to be interested in a topic and feel that they may contribute to its understanding, without any particular concern for whether there will be immediate, practical outcomes of the research findings. In this case, the goal of the research may be to learn more about a topic; for instance, what variables might be important to understand the experience of intramural sports. ***Applied research*** lies at the

other end of the continuum. In the social sciences, applied research typically refers to research that is conducted for some purpose beyond or in addition to a researcher's interest in contributing to understanding a topic.

Applied research is often client focused, meaning that the researcher is investigating a question posed by or of specific relevance to someone other than her or himself. As I describe later this chapter, in my first job after earning an undergraduate degree, the applied nature of the research I was hired to conduct lay in its solution- orientation. It wasn't simply to satisfy the curiosity of my employer, but was intended to be immediately applied to address certain, identifiable problems. What do you think the purpose of social scientific inquiry should be? Should social scientists conduct research for its own sake; only if it has some identifiable application; or, perhaps, for something in between?

One final consideration that social scientists must be aware of is the difference between qualitative and quantitative methods. ***Qualitative methods*** are ways of collecting data that yield results such as words or pictures. Some of the most common qualitative methods in the social sciences include field research, intensive interviews, and focus groups. ***Quantitative methods***, on the other hand, result in data that can be represented by and condensed into numbers. Survey research is probably the most common quantitative method in the applied human sciences, but methods such as content analysis and interviewing can also be conducted in a way that yields quantitative data. While qualitative methods aim to gain an in-depth understanding of a relatively small number of cases, quantitative methods offer less depth but more breadth because they typically focus on a much larger number of cases.

Sometimes these two methods are presented or discussed in a way that suggests they are somehow in opposition to one another. The qualitative/quantitative debate is fueled by researchers who may prefer one approach over another, either because their own research questions are better suited to one particular approach or because they happened to have been trained in one specific method. In this text, we'll operate from the perspective that qualitative and quantitative methods are complementary rather than competing. While these two methodological approaches certainly differ, the main point is that they simply have different goals, strengths, and weaknesses. That said, the focus of this text is quantitative research methods.

In sum, social scientists should be aware of the following considerations:

- There are several different ways that we know what we know, including informal observation, selective observation, overgeneralization, authority, and research methods.
- Research methods are a much more reliable source of knowledge than most of our other ways of knowing.
- A person's ontological perspective shapes her or his beliefs about the nature of reality, or what "is."
- A person's epistemological perspective shapes her or his beliefs about how we know what we know, and the best way(s) to uncover knowledge.

Conducting Ethical Research

In 1998, actor Jim Carey starred in the movie The Truman Show. At first glance, the film appears to depict a perfect sociological experiment. Just imagine the possibilities if we could control every aspect of a person's life, from how and where that person lives to where he or she works to whom he or she marries. Of course, keeping someone in a bubble, controlling every aspect of his or her life, and sitting back and watching would be highly unethical (not to mention illegal). However, the movie clearly inspires thoughts about the differences between researching humans versus inanimate objects. One of the most exciting—and most challenging—aspects of conducting research is the fact that (at least much of the time) our subjects are living human beings whose free will and human rights will always have an impact on what we are able to research and how we are able to conduct that research.

Unsurprisingly, research on human subjects is regulated much more heavily than research on nonhuman subjects. There are ethical considerations that all researchers must consider regardless of their research subject. As outlined in the Tri-Council Policy Statement, there are basic principles to follow, from which stem specific considerations when researching human subjects. We'll discuss those considerations following a brief outline of the TCPS2 "guiding principles."

The Tri-Council Policy Statement

The Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans is a definitive source of guidelines and best-practices when conducting research with human participants. It is a joint publication of the three main federal granting agencies in Canada, the Social Sciences and Humanities Research Council, the Natural Sciences and Engineering Research Council, and the Canadian Institutes of Health Research. The guidelines in the TCPS2 are grounded in the underlying value of respect for human dignity. This means that any research involving humans must be "sensitive to the inherent worth of all human beings and the respect and consideration that they are due" (TCPS2, 1998, p.6). This notion is further specified by the three core principles of "respect for persons," "concern for welfare," and "justice," each of which imply certain inherent rights of research participants:

- Respect for Persons
- Respect for free and informed consent
- Respect for vulnerable persons
- Concern for Welfare
- Respect for privacy and confidentiality

- Minimizing harm/maximizing benefits
- Balancing harm and benefits
- Justice
- Respect for Justice and Inclusiveness
- (Respect for Vulnerable Persons)

Knowing these principles in their general sense, as well as how such principles are applied to different issues and in different contexts, is essential to proper research conduct.

Specific Ethical Issues to Consider

As should be clear by now, conducting research on humans presents a number of unique ethical considerations. Human research subjects must be informed about and be given the opportunity to consent to their participation in research. Further, subjects' identities and the information they share should be protected by researchers. In this section, we'll take a look at a few specific topics that individual researchers must consider before embarking on research with human subjects.

Informed Consent

A norm of voluntary participation is presumed in all sociological research projects. In other words, we cannot force anyone to participate in our research without that person's knowledge or consent (so much for that Truman Show experiment). Researchers must therefore design procedures to obtain subjects' informed consent to participate in their research. Informed consent is defined as a subject's voluntary agreement to participate in research based on a full understanding of the research and of the possible risks and benefits involved. Although it sounds simple, ensuring that one has actually obtained informed consent is a much more complex process than you might initially presume.

The first requirement is that, in giving their informed consent, subjects may neither waive nor even appear to waive any of their legal rights. However, since social science research does not typically involve asking subjects to place themselves at risk of physical harm by, for example, taking untested drugs or consenting to new medical procedures, researchers do not often worry about potential liability associated with their research projects.

However, their research may involve other types of risks. For example, what if a researcher fails to sufficiently conceal the identity of a subject who admits to participating in a local swinger's club, enjoying a little sadomasochistic activity now and again or violating her marriage vows? While the

law may not have been broken in any of these cases, the subject's social standing, marriage, custody rights, or employment could be jeopardized were any of these tidbits to become public. This example might seem rather extreme, but the point remains: even social scientists conduct research that could come with some very real legal ramifications.

Beyond the legal issues, most institutional review boards (IRBs) require researchers to share some details about the purpose of the research, possible benefits of participation, and, most importantly, possible risks associated with participating in that research with their subjects. In addition, researchers must describe how they will protect subjects' identities, how and for how long any data collected will be stored, and whom to contact for additional information about the study or about subjects' rights. All this information is typically shared in an informed consent form that researchers provide to subjects. In some cases, subjects are asked to sign the consent form indicating that they have read it and fully understand its contents. In other cases, subjects are simply provided a copy of the consent form and researchers are responsible for making sure that subjects have read and understand the form before proceeding with any kind of data collection.

One last point to consider when preparing to obtain informed consent is that not all potential research subjects are considered equally competent or legally allowed to consent to participate in research. These subjects are sometimes referred to as members of vulnerable populations, people who may be at risk of experiencing undue influence or coercion.

The rules for consent are more stringent for vulnerable populations. For example, minors must have the consent of a legal guardian in order to participate in research. In some cases, the minors themselves are also asked to participate in the consent process by signing special, age-appropriate consent forms designed specifically for them. Prisoners and parolees also qualify as vulnerable populations. Concern about the vulnerability of these subjects comes from the very real possibility that prisoners and parolees could perceive that they will receive some highly desired reward, such as early release, if they participate in research. Another potential concern regarding vulnerable populations is that they may be underrepresented in research, and even denied potential benefits of participation in research, specifically because of concerns about their ability to consent. So on the one hand, researchers must take extra care to ensure that their procedures for obtaining consent from vulnerable populations are not coercive. And the procedures for receiving approval to conduct research on these groups may be more rigorous than that for non-vulnerable populations. On the other hand, researchers must work to avoid excluding members of vulnerable populations from participation simply on the grounds that they are vulnerable or that obtaining their consent may be more complex. While there is no easy solution to this double-edged sword, an awareness of the potential concerns associated with research on vulnerable populations is important for identifying whatever solution is most appropriate for a specific case.

Protection of Identities

As mentioned earlier, the informed consent process includes the requirement that researchers outline how they will protect the identities of subjects. This aspect of the process, however, is one of the most commonly misunderstood aspects of research.

In protecting subjects' identities, researchers typically promise to maintain either the anonymity or the confidentiality of their research subjects. Anonymity is the more stringent of the two. When a researcher promises anonymity to participants, not even the researcher is able to link participants' data with their identities. Anonymity may be impossible for some researchers to promise because several of the modes of data collection that social scientists employ, such as participant observation and face-to-face interviewing, require that researchers know the identities of their research participants. In these cases, a researcher should be able to at least promise confidentiality to participants. Offering confidentiality means that some identifying information on one's subjects is known and may be kept, but only the researcher can link participants with their data and he or she promises not to do so publicly. Sometimes it is not even possible to promise that a subject's confidentiality will be maintained. This can be the case if data are collected in public or in the presence of other research participants.

Protecting research participants' identities is not always a simple prospect, especially for those conducting research on stigmatized groups or illegal behaviors. Scott DeMuth learned that all too well when conducting his dissertation research on a group of animal rights activists. As a participant observer, DeMuth knew the identities of his research subjects. So when some of his research subjects vandalized facilities and removed animals from several research labs at the University of Iowa, a grand jury called on Mr. DeMuth to reveal the identities of the participants in the raid. When DeMuth refused to do so, he was jailed briefly and then charged with conspiracy to commit animal enterprise terrorism and cause damage to the animal enterprise (Jaschik, 2009).

Publicly, DeMuth's case raised many questions. What do social scientists owe the public? Is DeMuth, by protecting his research subjects, harming those whose labs were vandalized? Is he harming the taxpayers who funded those labs? Or is it more important that DeMuth emphasize what he owes his research subjects, who were told their identities would be protected? DeMuth's case also sparked controversy among academics, some of whom thought that as an academic himself, DeMuth should have been more sympathetic to the plight of the faculty and students who lost years of research as a result of the attack on their labs. Many others stood by DeMuth, arguing that the personal and academic freedom of scholars must be protected whether we support their research topics and subjects or not. What do you think? Should DeMuth have revealed the identities of his research subjects? Why or why not?

Minimising Harms

As illustrated above, there are a variety of ways in which our actions as researchers nmay put our participants at risk of harm. Whether such harm is professional, psychological, physical, or otherwise, it is our responsibility to do all that can be done to minimise them. While it may be that all risks cannot be eliminated without compromising the research, "should attempt to achieve the most favourable balance of risks and potential benefits" (TCPS2, p. 8) As well, we must be aware that harm may not only accrue to individual participants. Research, for instance, into marginalised neighborhoods or social groups may yield benefits, but there may also be the risk of further stigmatization or discrimination. In short, while much quantitative research may be deemed "minimal risk," we must be aware that risks of harm extend beyond the individual participants in our research.

Indeed, certain vulnerable populations deserve and require special consideration. The TCPS2 devotes a chapter to First Nations, Inuit and Metis peoples, and special protections are required for research with members of vulnerable populations, such as children, those of diminished mental capacity, and those under control of the state. The nuances involved in ethically researching these populations suggest that all researchers in Canada have more than a passing familiarity with the TCPS2.

Table 1.2 Ethical Consideration Concerning the Different Levels of Inquiry

Level of Inquiry	Focus	Key Ethical Questions
Micro	Individual	Does my research impinge on the individual's right to privacy?
		Could my research offend subjects in any way?
		Could my research cause emotional distress to any of my subjects?
		Has my own conduct been ethical throughout the research process?
Meso	Group	Does my research follow the ethical guidelines of my profession and discipline?
		Have I met my duty to those who funded my research?
Macro	Society	Does my research meet the societal expectations of social research?
		Have I met my social responsibilities as a researcher?

The Practice of Science Versus the Uses of Science

Research ethics has to do with both how research is conducted and how findings from that research are used and by whom. In this section, we'll consider research ethics from both angles.

Doing Science the Ethical Way

As you should now be aware, researchers must consider their own personal ethical principles in addition to following those of their institution, their discipline, and their community. We've already considered many of the ways that social scientists work to ensure the ethical practice of research, such as informing and protecting subjects. But the practice of ethical research doesn't end once subjects have been identified and data have been collected. Social scientists must also fully disclose their research procedures and findings. This means being honest about how research subjects were identified and recruited, how exactly data were collected and analyzed, and ultimately, what findings were reached.

If researchers fully disclose how they conducted their research, then those of us who use their work to build our own research projects, to create social policies, or to make decisions about our lives can have some level of confidence in the work. By sharing how research was conducted, a researcher helps assure readers that he or she has conducted legitimate research and didn't simply come to whatever conclusions he or she wanted to find. A description or presentation of research findings that is not accompanied by information about research methodology is missing some relevant information. Sometimes methodological details are left out because there isn't time or space to share them. This is often the case with news reports of research findings. Other times, there may be a more insidious reason that that important information isn't there. This may be the case if sharing methodological details would call the legitimacy of a study into question. As researchers, it is our ethical responsibility to fully disclose our research procedures. As consumers of research, it is our ethical responsibility to pay attention to such details. We'll discuss this more in the section "Using Science the Ethical Way."

The requirement of honesty comes not only from the principles of integrity and scientific responsibility but also out of the scientific principle of replication. Ideally, this means that one scientist could repeat another's study with relative ease. By replicating a study, we may become more (or less) confident in the original study's findings. Replication is far more difficult (perhaps impos-

sible) to achieve in the case of ethnographic studies that last months or years, but it nevertheless sets an important standard for all social scientific researchers: that we provide as much detail as possible about the processes by which we reach our conclusions.

Full disclosure also includes the need to be honest about a study's strengths and weaknesses, both with oneself and with others. Being aware of the strengths and weaknesses of one's own work can help a researcher make reasonable recommendations about the next steps other researchers might consider taking in their inquiries. Awareness and disclosure of a study's strengths and weaknesses can also help highlight the theoretical or policy implications of one's work. In addition, openness about strengths and weaknesses helps those reading the research better evaluate the work and decide for themselves how or whether to rely on its findings. Finally, openness about a study's sponsors is crucial. How can we effectively evaluate research without knowing who paid the bills?

The standard of replicability, along with openness about a study's strengths, weaknesses, and funders enable those who read the research to evaluate it fairly and completely. Knowledge of funding sources is often raised as an issue in medical research. Understandably, independent studies of new drugs may be more compelling to government agencies that regulate drug manufacture than studies touting the virtues of a new drug that happen to have been funded by the company who created that drug. But medical researchers aren't the only ones who need to be honest about their funding. If we know, for example, that a political think tank with ties to a particular party has funded some sociological research, we can take that knowledge into consideration when reviewing the study's findings and stated policy implications. Lastly, and related to this point, we must consider how, by whom, and for what purpose research may be used.

Using Science the Ethical Way

Science has many uses. By "use" I mean the ways that science is understood and applied (as opposed to the way it is conducted). Some use science to create laws and social policies; others use it to understand themselves and those around them. Some people rely on science to improve their life conditions or those of other people, while still others use it to improve their businesses or other undertakings. In each case, the most ethical way for us to use science is to educate ourselves about the design and purpose of any studies we may wish to use or apply, to recognize our limitations in terms of scientific and methodological knowledge and how those limitations may impact our understanding of research, and to apply the findings of scientific investigation only in cases or to populations for which they are actually relevant.

Social scientists who conduct research on behalf of organizations and agencies may face additional ethical questions about the use of their research, particularly when the organization for which an applied study is conducted controls the final report and the publicity it receives. Consider

a researcher hired by an organization to conduct leadership effectiveness research. While such research may not be inherently problematic, the potential conflict of interest between evaluation researchers and the employer being evaluated certainly exists. A similar conflict of interest might exist between independent researchers whose work is being funded by some government agency or private foundation.

So who decides what constitutes ethical conduct or use of research? Perhaps we all do. What qualifies as ethical research may shift over time and across cultures as individual researchers, disciplinary organizations, members of society, as well as regulatory entities such as institutional review boards, courts, and lawmakers all work to define the boundaries between ethical and unethical research.

Key Takeaways, Exercises and References

Key Takeaways

- Social sciences focus on aggregates and on patterns in society.
- Sometimes social science research is conducted for its own sake; other times it is focused on matters of public interest or on client-determined questions.
- Social scientists use both qualitative and quantitative methods. While different, these methods are often complementary.
- Whether we know it or not, our everyday lives are shaped by social scientific research.
- Understanding social scientific research methods can help us become more astute and more responsible consumers of information.
- Knowledge about social scientific research methods is useful for a variety of jobs or careers.
- Researchers are obliged to conduct their research ethically; in Canada, in conformity to the TCPS2 as well as any further institutional requirements.
- At the micro level, researchers should consider their own conduct and the rights of individual research participants.
- At the meso level, researchers should consider the expectations of their profession and of any organizations that may have funded their research.
- At the macro level, researchers should consider their duty to and the expectations of society with respect to social scientific research.
- Conducting research ethically requires that researchers be ethical not only in their data collection procedures but also in reporting their methods and findings.
- The ethical use of research requires an effort to understand research, an awareness of one's own limitations in terms of knowledge and understanding, and the honest application of research findings.
- What qualifies as ethical research is determined collectively by a number of individuals, organizations, and institutions and may change over time.

Exercises

- Think about a time in the past when you made a bad decision (e.g., wore the wrong shoes for hiking, dated the wrong person, chose not to study for an exam, got caught unprepared in a

rainstorm). What caused you to make this decision? How did any of the ways of knowing described previously contribute to your error-prone decision-making process? How might formal research methods help you overcome the possibility of committing such errors in the future?

- What should the purpose of social science be? Posit an argument in favor and against both applied and basic research.
- Scroll through a few popular websites or news sources. Pull out any examples you see of results from social science research being discussed. How much information about the research is provided? What questions do you have about the research? To what extent will the research shape your actions or beliefs? How, if at all, is your answer to that question based on your confidence in the research described?
- Scroll through the opinion pages of some news outlets. What kind of research would you like to see to accept or rebut the opinion provided? What questions might you have about any research that has been cited? In your opinion, has the opinion been adequately supported through evidence?
- Think of an instance when doing science ethically might conflict with using science ethically. Describe your example and how you, as a researcher, might proceed were you to find yourself in such a quandary.
- Using library and Internet resources, find three examples of funded research. Who were the funders in each case? How do the researchers inform readers about their funders? In what ways, if any, do you think each funder might influence the research? What questions, if any, do you have about the research after taking these potential influences into consideration?

Extras:

- The ASA website offers a case study of Rik Scarce's experience with protecting his data. You can read the case, and some thought- provoking questions about it, here: https://www.asanet.org/ethics/detail.cfm?id=Case99. What questions and concerns about conducting sociological research does Scarce's experience raise for you?
- The PBS series NOVA has an informative website and exercise on public opinion of the use of the Nazi experiment data. Go through the exercise at https://www.pbs.org/wgbh/nova/holocaust/experiments.html.

References

Adler, E. S., & Clark, R. (2011). *An invitation to social research: How it's done*. Belmont, CA: Wadsworth.

American Sociological Association. (1993). Case 99: A real case involving the protection of confidential data. Retrieved from https://www.asanet.org/ethics/detail.cfm?id=Case99

Berger, P. L. (1990). Nazi science: The Dachau hypothermia experiments. *New England Journal of Medicine*, 322, 1435–1440.

Bobbitt-Zeher, D., & Downey, D. B. (2010). Good for nothing? Number of siblings and friendship nominations among adolescents. Presented at the 2010 Annual Meeting of the American Sociological Association, Atlanta, GA.

Burawoy, M. (2005). 2004 presidential address: For public sociology. *American Sociological Review*, 70, 4–28.

Calhoun, C. (Ed.). (2007). *Sociology in America: A history*. Chicago, IL: University of Chicago Press.

Council of the Animals and Society Section of the American Sociological Association: Support for Scott DeMuth. (2009). Retrieved from https://davenportgrandjury.wordpress.com/solidarity-statements/council-animals- society-as

Elliott, W., Jung, H., Kim, K., & Chowa, G. (2010). A multi-group structural equation model (SEM) examining asset holding effects on educational attainment by race and gender. *Journal of Children & Poverty*, 16, 91–121.

Ellwood, D., & Kane, T. (2000). Who gets a university education? Family background and growing gaps in enrollment. In S. Danziger & J. Waldfogel (Eds.), *Securing the future* (pp. 283–324). New York, NY: Russell Sage Foundation.

Greene, V. W. (1992). Can scientists use information derived from the concentration camps? Ancient anwers to new questions. In A. L. Caplan (Ed.), *When medicine went mad: Bioethics and the Holocaust* (pp. 169–170). Totowa, NJ: Humana Press.

Henslin, J. M. (2006). *Essentials of sociology: A down-to-earth approach* (6th ed.). Boston, MA: Allyn and Bacon.

Jeffries, V. (Ed.). (2009). *Handbook of sociology*. Lanham, MD: Rowman & Littlefield.

Jenkins, P. J., & Kroll-Smith, S. (Eds.). (1996). *Witnessing for sociology: Social scientists in court*. Westport, CT: Praeger.

Mikkelson, B. (2005). *Grandma's cooking secret*. Retrieved from https://www.snopes.com/fact-check/grandmas-cooking-secret/

Moe, K. (1984). Should the Nazi research data be cited? *The Hastings Center Report*, 14, 5–7.

Pozos, R. S. (1992). Scientific inquiry and ethics: The Dachau data. In A. L. Caplan (Ed.), *When medicine went mad: Bioethics and the Holocaust* (p. 104). Totowa, NJ: Humana Press.

Sadker, M., & Sadker, D. (1994). *Failing at fairness: How America's schools cheat girls*. New York, NY: Maxwell Macmillan International.

Scarce v. United States, 5 F.3d 397, 399–400 (9th Cir. 1993).

CHAPTER 2: LINKING METHODS WITH THEORY

One of the hardest parts of research is moving beyond description to explanation. In our search for the explanation of social phenomena, we are guided by our paradigmatic frame and the types of approaches and theories that it gives rise to. In this chapter, we'll explore the connections between paradigms, social theories, and social scientific research methods. We'll also consider how one's analytic, paradigmatic, and theoretical perspective might shape or be shaped by her or his methodological choices. In short, we'll answer the question of what theory has to do with research methods.

LEARNING OBJECTIVES

- Describe a micro-level approach to research, and provide an example of a micro-level study.
- Describe a meso-level approach to research, and provide an example of a meso-level study.
- Describe a macro-level approach to research, and provide an example of a macro-level study.
- Define paradigm, and describe the significance of paradigms.
- Identify and describe the four predominant paradigms found in the social sciences.
- Define theory.
- Describe the role that theory plays in social scientific inquiry.
- Understand how different levels of analysis and different approaches, such as inductive and deductive, can shape the way that a topic is investigated.
- Describe the inductive approach to research, and provide examples of inductive research.
- Describe the deductive approach to research, and provide examples of deductive research.
- Describe the ways that inductive and deductive approaches may be complementary.

Micro, Meso, and Macro Approaches

Before we discuss the more specific details of paradigms and theories, let's look broadly at three possible levels of inquiry on which social scientific investigations might be based. These three levels demonstrate that while social scientists share some common beliefs about the value of investigating and understanding human interaction, at what level they investigate that interaction will vary.

At the micro level, social scientists examine the smallest levels of interaction; even in some cases, just "the self" alone. Micro-level analyses might include one-on-one interactions between couples or friends. Or perhaps a researcher is interested in how a person's perception of self is influenced by his or her social context. In each of these cases, the level of inquiry is micro. Meso-level analyses concern groups, and social scientists who conduct meso-level research might study, for instance, how norms of workplace behavior vary across professions or how children's sporting clubs are organized. At the macro level, social scientists examine social structures and institutions. Research at the macro level examines large-scale patterns. A study of globalization that examines immigration patters between nations would be an example of a macro-level study.

Social Science at Three Different Levels

Let's take a closer look at some specific examples of research to better understand each of the three levels of inquiry described previously. Some topics are best suited to be examined at one particular level, while other topics can be studied at each of the three different levels. The particular level of inquiry might shape a social scientist's questions about the topic, or a social scientist might view the topic from different angles depending on the level of inquiry being employed.

First, let's consider some examples of different topics that are best suited to a particular level of inquiry. Work by Stephen Marks offers an excellent example of research at the ***micro-level***. In one study, Marks and Shelley MacDermid (1996) draw from prior micro-level theories to empirically study how people balance their roles and identities. In this study, the researchers found that people who experience balance across their multiple roles and activities report lower levels of depression and higher levels of self-esteem and well-being than their less- balanced counterparts. In another study, Marks and colleagues examined the conditions under which husbands and wives feel the most balance across their many roles. They found that different factors are important for different genders. For women, having more paid work hours and more couple time were among the most important factors. For men, having leisure time with their nuclear families was important, and role balance decreased as work hours increased (Marks, Huston, Johnson, & MacDermid, 2001). Both of these studies fall within the category of micro-level analysis.

At the ***meso-level***, social scientists tend to study the experiences of groups and the interactions between groups. In a recent book based on their research with Somali immigrants, Kim Huisman and colleagues (Huisman, Hough, Langellier, & Toner, 2011) examine the interactions between Somalis and Americans in Maine. These researchers found that stereotypes about refugees being unable or unwilling to assimilate and being overly dependent on local social systems are unsubstantiated. In a much different study of group-level interactions, Michael Messner (2009) conducted research on children's sports leagues. Messner studied interactions among parent volunteers, among youth participants, and between league organizers and parents and found that gender boundaries and hierarchies are perpetuated by the adults who run such leagues. These two studies, while very different in their specific points of focus, have in common their meso-level focus.

Social scientists who conduct ***macro-level*** research study interactions at the broadest level, such as interactions between nations or comparisons across nations. Frank, Camp, & Boutcher (2010) provide one example of macro-level research. These researchers examined worldwide changes over time in laws regulating sex. By comparing laws across a number of countries over a period of many years (1945–2005), Frank learned that laws regulating rape, adultery, sodomy, and child sexual abuse shifted in focus from protecting larger entities, such as families, to protecting individuals. In another macro-level study, Leah Ruppanner (2010) [6] studied how national levels of gender equality in 25 different countries affect couples' divisions of housework. Ruppanner found, among other patterns, that as women's parliamentary representation increases, so, too, does men's participation in housework.

While it is true that some topics lend themselves to a particular level of inquiry, there are many topics that could be studied from any of the three levels. The choice depends on the specific interest of the researcher, the approach he or she would like to take, and the sorts of questions he or she wants to be able to answer about the topic. Let's look at an example. Gang activity has been a topic of interest to social scientists for many years and has been studied from each of the levels of inquiry described here. At the micro level, social scientists might study the inner workings of a specific gang, communication styles, and what everyday life is like for gang members. Though not a scholarly account, one example of a micro-level analysis of gang activity can be found in Sanyika Shakur's 1993 autobiography, Monster. In his book, Shakur describes his former day-to-day life as a member of the Crips in south-central Los Angeles. Shakur's recounting of experiences highlights micro-level interactions between himself, fellow Crips members, and other gangs.

At the meso-level, social scientists are likely to examine interactions between gangs or perhaps how different branches of the same gang vary from one area to the next. At the macro level, we could compare the impact of gang activity across communities or examine the economic impact of gangs on nations. Excellent examples of gang research at all three levels of analysis can be found in the Journal of Gang Research published by the National Gang Crime Research Center (NGCRC). Sudir Venkatesh's study (2008), Gang Leader for a Day, is an example of research on gangs that utilizes all three levels of analysis. Venkatesh conducted participant observation with a gang in

Chicago. He learned about the everyday lives of gang members (micro) and how the gang he studied interacted with and fit within the landscape of other gang "franchises" (meso). In addition, Venkatesh described the impact of the gang on the broader community and economy (macro).

Paradigms, Theories, and How They Shape a Researcher's Approach

The terms paradigm and theory are often used interchangeably in social science, although social scientists do not always agree whether these are identical or distinct concepts. In this text, we will make a slight distinction between the two ideas because thinking about each concept as analytically distinct provides a useful framework for understanding the connections between research methods and social scientific ways of thinking.

Paradigms in Social Science

For our purposes, we'll define paradigm as an analytic lens, a way of viewing the world and a framework from which to understand the human experience (Kuhn, 1962). It can be difficult to fully grasp the idea of paradigmatic assumptions because we are very ingrained in our own, personal everyday way of thinking. For example, let's look at people's views on abortion. To some, abortion is a medical procedure that should be undertaken at the discretion of each individual woman who might experience an unwanted pregnancy. To others, abortion is murder and members of society should collectively have the right to decide when, if at all, abortion should be undertaken. Chances are, if you have an opinion about this topic you are pretty certain about the veracity of your perspective. Then again, the person who sits next to you in class may have a very different opinion and yet be equally confident about the truth of his or her perspective. Which of you is correct? You are each operating under a set of assumptions about the way the world does—or at least should—work. Perhaps your assumptions come from your particular political perspective, which helps shape your view on a variety of social issues, or perhaps your assumptions are based on what you learned from your parents or in church. In any case, there is a paradigm that shapes your stance on the issue.

In Chapter 1, we discussed the various ways that we know what we know. Paradigms are a way of framing what we know, what we can know, and how we can know it. In social science, there are several predominant paradigms, each with its own unique ontological and epistemological perspective. Let's look at four of the most common social scientific paradigms that might guide you as you begin to think about conducting research.

The first paradigm we'll consider, called ***positivism***, is probably the framework that comes to mind for many of you when you think of science. Positivism is guided by the principles of objectivity, knowability, and deductive logic. Deductive logic is discussed in more detail in the section that follows. Auguste Comte, whom you might recall from your introduction to sociology class as the

person who coined the term sociology, argued that sociology should be a positivist science (Ritzer & Goodman, 2004). [2] The positivist framework operates from the assumption that society can and should be studied empirically and scientifically. Positivism also calls for value-free science, one in which researchers aim to abandon their biases and values in a quest for objective, empirical, and knowable truth.

Another predominant paradigm in the social sciences is ***social constructionism***. While positivists seek "the truth," the social constructionist framework posits that "truth" is a varying, socially constructed, and ever- changing notion. This is because we, according to this paradigm, create reality ourselves (as opposed to it simply existing and us working to discover it) through our interactions and our interpretations of those interactions. Key to the social constructionist perspective is the idea that social context and interaction frame our realities.

Researchers operating within this framework take keen interest in how people come to socially agree, or disagree, about what is real and true. Consideration of how meanings of different hand gestures vary across different regions of the world aptly demonstrates that meanings are constructed socially and collectively. Think about what it means to you when you see a person raise his or her middle finger. We probably all know that person isn't very happy (nor is the person to whom the finger is being directed). In some societies, it is another gesture, the thumbs up, that raises eyebrows. While the thumbs up may have a particular meaning in our culture, that meaning is not shared across cultures (Wong, 2007). [4]

It would be a mistake to think of the social constructionist perspective as only individualistic. While individuals may construct their own realities, groups—from a small one such as a married couple to large ones such as nations—often agree on notions of what is true and what "is." In other words, the meanings that we construct have power beyond the individual people who create them. Therefore, the ways that people work to change such meanings is of as much interest to social constructionists as how they were created in the first place.

A third paradigm is the ***critical paradigm***. At its core, the critical paradigm is focused on power, inequality, and social change. Although some rather diverse perspectives are included here, the critical paradigm, in general, includes ideas developed by early social theorists, such as Max Horkheimer (Calhoun, Gerteis, Moody, Pfaff, & Virk), and later works developed by feminist scholars, such as Nancy Fraser (1989). Unlike the positivist paradigm, the critical paradigm posits that social science can never be truly objective or value-free. Further, this paradigm operates from the perspective that scientific investigation should be conducted with the express goal of social change in mind.

Finally, ***postmodernism*** is a paradigm that challenges almost every way of knowing that many social scientists take for granted (Best & Kellner, 1991). While positivists claim that there is an objective, knowable truth, postmodernists would say that there is not. While social constructionists may argue that truth is in the eye of the beholder (or in the eye of the group that agrees on it), postmodernists may claim that we can never really know such truth because, in the studying and reporting of others' truths, the researcher stamps her or his own truth on the investigation. Finally, while the

critical paradigm may argue that power, inequality, and change shape reality and truth, a postmodernist may in turn ask, whose power, whose inequality, whose change, whose reality, and whose truth? As you might imagine, the postmodernist paradigm poses quite a challenge for social scientific researchers. How does one study something that may or may not be real or that is only real in your current and unique experience of it? This fascinating question is worth pondering as you begin to think about conducting your own research, though a deep dive into postmodernism can probably wait until grad school. Table 2.1 "Social Scientific Paradigms" summarizes each of the paradigms discussed here.

Table 2.1 Social Scientific Paradigms

Paradigm	Emphasis	Assumption
Positivism	Objectivity, knowability, and deductive logic	Society can and should be studied empirically and scientifically.
Social constructionism	Truth as varying, socially constructed, and ever-changing	Reality is created collectively and that social context and interaction frame our realities.
Critical	Power, inequality, and social change	Social science can never be truly value-free and should be conducted with the express goal of social change in mind.
Postmodernism	Inherent problems with previous paradigms.	Truth in any form may or may not be knowable.

Social Science Theories

Much like paradigms, theories provide a way of looking at the world and of understanding human interaction. Like paradigms, theories can be sweeping in their coverage. Some sociological theories, for example, aim to explain the very existence and continuation of society as we know it. Unlike paradigms, however, theories might be narrower in focus, perhaps just aiming to understand one particular phenomenon, without attempting to tackle a broader level of explanation. In a nutshell, theory might be thought of as a way of explanation or as "an explanatory statement that fits the evidence" (Quammen, 2004). At their core, theories can be used to provide explanations of any number or variety of phenomena. They help us answer the "why" questions we often have about the patterns we observe in social life. Theories also often help us answer our "how" questions. While paradigms may point us in a particular direction with respect to our "why" questions, theories more specifically map out the explanation, or the "how," behind the "why."

A further complication is the existence of certain "grand" theories, those that—like paradigms—shape the way we approach and try to answer questions. For instance, introductory sociology textbooks typically teach students about "the big three" sociological theories—structural functionalism, conflict theory, and symbolic interactionism (Barkan, 2011; Henslin, 2010). Structural functionalists focus on the interrelations between various parts of society and how each part works with the others to make society function in the way that it does. Conflict theorists are interested in questions of power and who wins and who loses based on the way that society is organized. Finally, symbolic interactionists focus on how meaning is created and negotiated though meaningful (i.e., symbolic) interactions. Just as researchers might examine the same topic from different levels of inquiry, so, too, could they investigate the same topic from different theoretical perspectives. In this case, even their research questions could be the same, but the way they make sense of whatever phenomenon it is they are investigating will be shaped in large part by the theoretical assumptions that lie behind their investigation.

Table 2.2 "Sociological Theories and the Study of Sport" summarizes the major points of focus for each of major three theories and outlines how a researcher might approach the study of the same topic, in this case the study of sport, from each of the three perspectives.

Table 2.2 Sociological Theories and The Study of Sport

Paradigm	Focuses on	A study of sport might examine
Structural functionalism	Interrelations between parts of society; how parts work together	Positive, negative, intended, and unintended consequences of professional sport leagues
Conflict theory	Who wins and who loses based on the way that society is organized	Issues of power in sport such as differences in access to and participation in sport
Symbolic interactionism	How meaning is created and negotiated though interactions	How the rules of sport of are constructed, taught, and learned

Subordinate to these grand perspectives, there are many other theories that aim to explain more specific types of interactions. For example, within the study of sexual harassment, different theories posit different explanations for why harassment occurs. One theory, first developed by criminologists, is called routine activities theory. It posits that sexual harassment is most likely to occur when a workplace lacks unified groups and when potentially vulnerable targets and motivated offenders are both present (DeCoster, Estes, & Mueller, 1999). Other theories of sexual harassment, called relational theories, suggest that a person's relationships, such as their marriages or friendships, are the key to understanding why and how workplace sexual harassment occurs and how people will respond to it when it does occur (Morgan, 1999). Relational theories focus on the power that different social relationships provide (e.g., married people who have supportive partners at home might be more likely than those who lack support at home to report sexual harassment when

it occurs). Finally, feminist theories of sexual harassment take a different stance. These theories posit that the way our current gender system is organized, where those who are the most masculine have the most power, best explains why and how workplace sexual harassment occurs (MacKinnon, 1979). [13] As you might imagine, which theory a researcher applies to examine the topic of sexual harassment will shape the questions the researcher asks about harassment. It will also shape the explanations the researcher provides for why harassment occurs.

Inductive or Deductive? Two Different Approaches

Theories structure and inform sociological research. So, too, does research structure and inform theory. The reciprocal relationship between theory and research often becomes evident to students new to these topics when they consider the relationships between theory and research in inductive and deductive approaches to research. In both cases, theory is crucial. But the relationship between theory and research differs for each approach.

Inductive and deductive approaches to research are quite different, but they can also be complementary. Let's start by looking at each one and how they differ from one another. Then we'll move on to thinking about how they complement one another.

Inductive Approaches and Some Examples

In an inductive approach to research, a researcher begins by collecting data that is relevant to his or her topic of interest. Once a substantial amount of data have been collected, the researcher will then take a breather from data collection, stepping back to get a bird's eye view of her data. At this stage, the researcher looks for patterns in the data, working to develop a theory that could explain those patterns. Thus when researchers take an inductive approach, they start with a set of observations and then they move from those particular experiences to a more general set of propositions about those experiences. In other words, they move from data to theory, or from the specific to the general.

There are many good examples of inductive research, but we'll look at just a few here. One fascinating recent study in which the researchers took an inductive approach was Katherine Allen, Christine Kaestle, and Abbie Goldberg's study (2011) of how boys and young men learn about menstruation. To understand this process, Allen and her colleagues analyzed the written narratives of 23 young men in which the men described how they learned about menstruation, what they thought of it when they first learned about it, and what they think of it now. By looking for patterns across all 23 men's narratives, the researchers were able to develop a general theory of how

boys and young men learn about this aspect of girls' and women's biology. They conclude that sisters play an important role in boys' early understanding of menstruation, that menstruation makes boys feel somewhat separated from girls, and that as they enter young adulthood and form romantic relationships, young men develop more mature attitudes about menstruation.

In another inductive study, Kristin Ferguson and colleagues (Ferguson, Kim, & McCoy, 2011) analyzed empirical data to better understand how best to meet the needs of young people who are homeless. The authors analyzed data from focus groups with 20 young people at a homeless shelter. From these data they developed a set of recommendations for those interested in applied interventions that serve homeless youth. The researchers also developed hypotheses for people who might wish to conduct further investigation of the topic. Though Ferguson and her colleagues did not test the hypotheses that they developed from their analysis, their study ends where most deductive investigations begin: with a set of testable hypotheses.

Deductive Approaches and Some Examples

Researchers taking a deductive approach take the steps described earlier for inductive research and reverse their order. They start with a social theory that they find compelling and then test its implications with data. That is, they move from a more general level to a more specific one. A deductive approach to research is the one that people typically associate with scientific investigation. The researcher studies what others have done, reads existing theories of whatever phenomenon he or she is studying, and then tests hypotheses that emerge from those theories.

While not all researchers follow a deductive approach, as you have seen in the preceding discussion, many do, and there are a number of excellent recent examples of deductive research. We'll take a look at a couple of those next.

In a study of US law enforcement responses to hate crimes, Ryan King and colleagues (King, Messner, & Baller, 2009) hypothesized that law enforcement's response would be less vigorous in areas of the country that had a stronger history of racial violence. The authors developed their hypothesis from their reading of prior research and theories on the topic. , they tested the hypothesis by analyzing data on states' lynching histories and hate crime responses. Overall, the authors found support for their hypothesis.

In another deductive study, Melissa Milkie and Catharine Warner (2011) studied the effects of different classroom environments on first graders' mental health. Based on prior research and theory, Milkie and Warner hypothesized that negative classroom features, such as a lack of basic supplies and even heat, would be associated with emotional and behavioral problems in children. The researchers found support for their hypothesis, demonstrating that policymakers should probably be paying more attention to the mental health outcomes of children's school experiences, just as they track academic outcomes (American Sociological Association, 2011).

Complementary Approaches?

While inductive and deductive approaches to research seem quite different, they can actually be rather complementary. In some cases, researchers will plan for their research to include multiple components, one inductive and the other deductive. In other cases, a researcher might begin a study with the plan to only conduct either inductive or deductive research, but then he or she discovers along the way that the other approach is needed to help illuminate findings. Here is an example of each such case.

In the case of my collaborative research on the experience of short-term unemployment, from the outset the decision was made to pursue both a deductive and an inductive approach in our work. Respondents completed a quantitative survey as well as periodic, shorter questionnaires as part of experience sampling methodology, and qualitative interviews were conducted with participants. The survey data were well suited to a deductive approach; we could analyze those data to test hypotheses that were generated based on theories of unemployment. The interview data were well suited to an inductive approach; we looked for patterns across the interviews and then tried to make sense of those patterns by theorizing about them (Havitz, Morden, & Samdahl, 2004).

For instance, for our initial analysis we utilized an inductive approach to search across cases to begin to formulate broader understandings of the experience of unemployment, as well as an understanding of any specific similarities across cases. As a result of living with this qualitative data, we came to realize that those who we labelled "adult children living at home" (ACH) expressed far greater dissatisfaction with their current bout of unemployment than other unemployed people in our sample. We knew from the literature that being homebound was a potent driver of dissatisfaction during unemployment, so we hypothesized that ACH were more homebound than other unemployed people. We then tested our hypotheses by analyzing the experience sampling data. In assessing this hypothesis with the quantitative data, we did find that ACH spent significantly more time at home than other unemployed people (Havits, Morden, & Samdahl, 2004). It is important to note that we did not initially hypothesize about what we might find but, instead, inductively analyzed the interview data, looking for patterns that might tell us something about commonalities and differences in the experience of unemployment. Overall, our desire to understand unemployment experiences fully—in terms of their objective experiences, their perceptions of those experiences, and their stories of their experiences—led us to adopt both deductive and inductive approaches in the work.

Researchers may not always set out to employ both approaches in their work but sometimes find that their use of one approach leads them to the other. One such example is described eloquently in Schutt's (2006), *Investigating the Social World*. As Schutt describes, researchers Sherman and Berk (1984) conducted an experiment to test two competing theories of the effects of punishment on deterring deviance (in this case, domestic violence). Specifically, Sherman and Berk hypothesized that deterrence theory would provide a better explanation of the effects of arresting accused bat-

terers than labeling theory. Deterrence theory predicts that arresting an accused spouse batterer will reduce future incidents of violence. Conversely, labeling theory predicts that arresting accused spouse batterers will increase future incidents.

Sherman and Berk found, after conducting an experiment with the help of local police in one city, that arrest did in fact deter future incidents of violence, thus supporting their hypothesis that deterrence theory would better predict the effect of arrest. After conducting this research, they and other researchers went on to conduct similar experiments in six additional cities (Berk, Campbell, Klap, & Western, 1992; Pate & Hamilton, 1992; Sherman & Smith, 1992). Results from these follow-up studies were mixed. In some cases, arrest deterred future incidents of violence. In other cases, it did not. This left the researchers with new data that they needed to explain. The researchers therefore took an inductive approach in an effort to make sense of their latest empirical observations. The new studies revealed that arrest seemed to have a deterrent effect for those who were married and employed but that it led to increased offenses for those who were unmarried and unemployed. Researchers thus turned to control theory, which predicts that having some stake in conformity through the social ties provided by marriage and employment, as the better explanation.

What the Sherman and Berk research, along with the follow-up studies, shows us is that we might start with a deductive approach to research, but then, if confronted by new data that we must make sense of, we may move to an inductive approach.

Revisiting an Earlier Question

At the beginning of this chapter I asked, what's theory got to do with it? Perhaps at the time, you weren't entirely sure, but I hope you now have some ideas about how you might answer the question. Just in case, let's review the ways that theories are relevant to social scientific research methods.

Theories, paradigms, levels of analysis, and the order in which one proceeds in the research process all play an important role in shaping what we ask about the social world, how we ask it, and in some cases, even what we are likely to find. A micro-level study of gangs will look much different than a macro-level study of gangs. In some cases you could apply multiple levels of analysis to your investigation, but doing so isn't always practical or feasible. Therefore, understanding the different levels of analysis and being aware of which level you happen to be employing is crucial. One's theoretical perspective will also shape a study. In particular, the theory invoked will likely shape not only the way a question about a topic is asked but also which topic gets investigated in the first place. Further, if you find yourself especially committed to one paradigm over another, the possible answers you are likely to see to the questions that you pose are limited.

This does not mean that social science is biased or corrupt. At the same time, we humans can never claim to be entirely value free. Social constructionists and postmodernists might point out that bias is always a part of research to at least some degree. Our job as researchers is to recognize and address our biases as part of the research process, if an imperfect part. We all use particular approaches, be they theories, levels of analysis, or temporal processes, to frame and conduct our work. Understanding those frames and approaches is crucial not only for successfully embarking upon and completing any research-based investigation but also for responsibly reading and understanding others' work. So what's theory got to do with it? Just about everything.

Key Takeaways, Exercises and References

Key Takeaways

- Social science research can occur at any of three analytical levels: micro, meso, or macro.
- Some topics lend themselves to one particular analytical level while others could be studied from any, or all, of the three levels of analysis.
- Different levels of analysis lead to different points of focus on any given topic.
- Paradigms shape our everyday view of the world.
- Social scientists use theory to help frame their research questions and to help them make sense of the answers to those questions.
- Some theories are rather sweeping in their coverage and attempt to explain, broadly, how and why societies (or individual psyches, or national economies, etc.) are organized in particular ways.
- The theory being invoked, and the paradigm from which a researcher frames his or her work, can shape not only the questions asked but also the answers discovered.
- Whether a researcher takes an inductive or deductive approach will determine the process by which he or she attempts to answer his or her research question.
- The inductive approach involves beginning with a set of empirical observations, seeking patterns in those observations, and then theorizing about those patterns.
- The deductive approach involves beginning with a theory, developing hypotheses from that theory, and then collecting and analyzing data to test those hypotheses.
- Inductive and deductive approaches to research can be employed together for a more complete understanding of the topic that a researcher is studying.
- Though researchers don't always set out to use both inductive and deductive strategies in their work, they sometimes find that new questions arise in the course of an investigation that can best be answered by employing both approaches.

Exercises

- Of the four paradigms described, which do you find most compelling? Why?
- Think about a topic that you'd like to study. From what analytical level do you think it makes sense to study your topic? Why?

- Find an example of published research that examines a single topic from each of the three analytical levels. Describe how the researcher employs each of the three levels in her or his analysis.
- For a hilarious example of logic gone awry, check out the following clip from Monty Python and Holy Grail: https://www.youtube.com/watch?feature=player_embedded&v=yp_l5ntikaU
- Do the townspeople take an inductive or deductive approach to determine whether the woman in question is a witch? What are some of the different sources of knowledge (recall Chapter 1 "Introduction") they rely on?
- Think about how you could approach a study of the relationship between gender and driving over the speed limit. How could you learn about this relationship using an inductive approach? What would a study of the same relationship look like if examined using a deductive approach? Try the same thing with any topic of your choice. How might you study the topic inductively? Deductively?

References

Allen, K. R., Kaestle, C. E., & Goldberg, A. E. (2011). More than just a punctuation mark: How boys and young men learn about menstruation. Journal of Family Issues, 32, 129–156.

Berk, R., Campbell, A., Klap, R., & Western, B. (1992). The deterrent effect of arrest in incidents of domestic violence: A Bayesian analysis of four field experiments. American Sociological Review, 57, 698–708;

Best, S., & Kellner, D. (1991). Postmodern theory: Critical interrogations. New York, NY: Guilford.

Blackstone, A., Houle, J., & Uggen, C. "At the time I thought it was great": Age, experience, and workers' perceptions of sexual harassment. Presented at the 2006 meetings of the American Sociological Association. Currently under review.

Calhoun, C., Gerteis, J., Moody, J., Pfaff, S., & Virk, I. (Eds.). (2007). Classical sociological theory (2nd ed.). Malden, MA: Blackwell.

DeCoster, S., Estes, S. B., & Mueller, C. W. (1999). Routine activities and sexual harassment in the workplace. Work and Occupations, 26, 21–49.

Ferguson, K. M., Kim, M. A., & McCoy, S. (2011). Enhancing empowerment and leadership among homeless youth in agency and community settings: A grounded theory approach. Child and Adolescent Social Work Journal, 28, 1–22.

Frank, D., Camp, B., & Boutcher, S. (2010). Worldwide trends in the criminal regulation of sex, 1945–2005. American Sociological Review, 75, 867–893.

Fraser, N. (1989). Unruly practices: Power, discourse, and gender in cotemporary social theory. Minneapolis, MN: University of Minnesota Press.

Havitz, M. E., Morden, P. A. & Samdahl, D. (2004). The diverse experiences of unemployed adults: Implications for work, leisure, and well-being. Waterloo, ON: WLU P

Huisman, K. A., Hough, M., Langellier, K. M., & Toner, C. N. (2011). Somalis in Maine: Crossing cultural currents. New York, NY: Random House.

King, R. D., Messner, S. F., & Baller, R. D. (2009). Contemporary hate crimes, law enforcement, and the legacy of racial violence. American Sociological Review, 74, 291–315.

Kuhn, T. (1962). The structure of scientific revolutions. Chicago, IL: University of Chicago Press.

MacKinnon, C. 1979. Sexual harassment of working women: A case of sex discrimination. New Haven, CT: Yale University Press.

Marks, S. R., & MacDermid, S. M. (1996). Multiple roles and the self: A theory of role balance.Journal of Marriage and the Family, 58, 417–432.

Marks, S. R., Huston, T. L., Johnson, E. M., & MacDermid, S. M. (2001). Role balance among white married couples. Journal of Marriage and the Family, 63, 1083–1098.

Messner, M. A. (2009). It's all for the kids: Gender, families, and youth sports. Berkeley, CA: University of California Press.

Milkie, M. A., & Warner, C. H. (2011). Classroom learning environments and the mental health of first grade children. Journal of Health and Social Behavior, 52, 4–22.

Morgan, P. A. (1999). Risking relationships: Understanding the litigation choices of sexually harassed women. The Law and Society Review, 33, 201–226.

Pate, A., & Hamilton, E. (1992). Formal and informal deterrents to domestic violence: The Dade county spouse assault experiment. American Sociological Review, 57, 691–697;

Quammen, D. (2004, November). Was Darwin wrong? National Geographic, pp. 2–35.

Ritzer, G., & Goodman, D. J. (2004). Classical sociological theory (4th ed.). New York, NY: McGraw-Hill.

Ruppanner, L. E. (2010). Cross-national reports of housework: An investigation of the gender empowerment measure. Social Science Research, 39, 963–975.

Shakur, S. (1993). Monster: The autobiography of an L.A. gang member. New York, NY: Atlantic Monthly Press.

Sherman, L. W., & Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. American Sociological Review, 49, 261–272.

Sherman, L., & Smith, D. (1992). Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. American Sociological Review, 57, 680–690.

The American Sociological Association wrote a press release on Milkie and Warner's findings: American Sociological Association. (2011). Study: Negative classroom environment adversely affects children's mental health. Retrieved from https://asanet.org/press/Negative_Classroom_Environment_Adversely_Affects_Childs_Mental_Health.cfm

The Journal of Gang Research is the official publication of the National Gang Crime Research Center (NGCRC). You can learn more about the NGCRC and the journal at https://www.ngcrc.com.

Uggen, C., & Blackstone, A. (2004). Sexual harassment as a gendered expression of power.American Sociological Review, 69, 64–92.

Venkatesh, S. (2008). Gang leader for a day: A rogue social scientist takes to the streets. New York, NY: Penguin Group.

CHAPTER 3: DEVELOPING A RESEARCH PROJECT

Do you like to watch movies? Do you have a pet that you care about? Do you wonder what you and your peers might do with your degrees once you've finished university? Do you wonder how many people on your campus have heard of the Oka Crisis, how many know that Justin Trudeau is our Prime Minister, or how many know that tuition covers less than 20% of university operating expenses in Quebec? Have you ever felt that you were treated differently at work because of your gender, ethnicity, or mother tongue? If you answered yes to any of these questions, then you have just the sort of intellectual curiosity required to conduct a research project.

LEARNING OBJECTIVES

Define starting where you are, and describe how it works.

- Identify and describe two overarching questions researchers should ask themselves about where they already are.
- Define empirical questions, and provide an example.
- Define ethical questions, and provide an example.
- Understand and describe the differences among exploratory, descriptive, and explanatory research.
- Define and provide an example of idiographic research.
- Define and provide an example of nomothetic research.
- Identify circumstances under which research would be defined as applied and compare those to circumstances under which research would be defined as basic.
- Identify and explain the five key features of a good research question.
- Explain why it is important for social scientists to be focused when designing a research question.
- Identify the differences between and provide examples of strong and weak research questions.

- Describe the role of causality in quantiative research
- Identify, define, and describe each of the three main criteria for causality.
- Describe the difference between and provide examples of independent and dependent variables.

- Define units of analysis and units of observation, and describe the two common errors people make when they confuse the two.
- Define hypothesis, be able to state a clear hypothesis, and discuss the respective roles of quantitative and qualitative research when it comes to hypotheses.

Starting Where You Already Are

The preceding questions are all real questions that real students have asked in a research methods class just like the one that you are currently taking. In some cases, these students knew they had a keen interest in a topic before beginning their research methods class. For example, Matt, a Recreation and Leisure Studies major, started off with an interest in a focused topic. He had begun to worry about what he would do with his degree when he graduated, and so he designed a project to learn more about what other RLS majors did and planned to do.

In other cases, students did not start out with a specific interest linked to their academic pursuits, but these students, too, were able to identify research topics worthy of investigation. These students knew, for example, how they enjoyed spending their free time. Perhaps at first these students didn't realize that they could identify and answer a research question about their hobbies, but they certainly learned that they could once they had done a little brainstorming. For example, Aisha enjoyed reading about and watching movies, so she conducted a project on the relationship between movie reviews and movie success. Sarah, who enjoyed spending time with her pet cat, designed a project to learn more about therapeutic animal–human relationships.

Even students who claimed to have "absolutely no interests whatsoever" usually discovered that they could come up with a research question simply by stepping back, taking a bird's eye view of their daily lives, and identifying some interesting patterns there. This was the case for Allison, who made some remarkable discoveries about her restaurant job, where she had applied to work as a cook but was hired to work as a waitress. When Allison realized that all the servers at the restaurant were women and all the cooks were men, she began to wonder whether employees had been assigned different roles based on their gender identities. Allison's epiphany led her to investigate how jobs and workplace stereotypes are gendered. Like Allison, Alejandra also struggled to identify a research topic. Her academic experiences had not inspired any specific research interests, and when asked about hobbies, she claimed to have none. When asked what really annoys her, it occurred to Alejandra that she resented the amount of time her friends spent watch and discussing the reality television show The Bachelorette. This realization led Teresa to her own aha moment: She would investigate who watches reality television and why.

Whether it was thinking about a question they'd had for some time, identifying a question about their own interests and hobbies, or taking a look at patterns in their everyday life, every student in these research methods classes managed two things: to identify a research question that was of interest to them and to foresee what data was needed to help answer that question. In this chapter, we'll focus on how to identify possible topics for study, how to make your topic appropriate for applied human sciences, how to phrase your interest as a research question, and how to get started once you have identified that question. In later chapters, we'll learn more about how to actually answer the questions you will have developed by the time you finish this chapter.

Once you have identified where you already are, there are two overarching questions you need to ask yourself: how do you feel about where you already are; and, what do you know about where you already are?

How Do You Feel About Where You Already Are?

Once you have figured out where you already are, your next task is to ask yourself some important questions about the interest you've identified. Your answers to these questions will help you decide whether your topic is one that will really work for a research project.

Whether you begin by already having an interest in some topic or you decide you want to study something related to one of your hobbies or your everyday experiences, chances are good that you already have some opinions about your topic. As such, there are a few questions you should ask yourself to determine whether you should try to turn this topic into a research project.

Start by asking yourself how you feel about your topic. Be totally honest, and ask yourself whether you believe your perspective is the only valid one. Perhaps yours isn't the only perspective, but do you believe it is the wisest one?

The most practical one? How do you feel about other perspectives on this topic? If you feel so strongly that certain findings would upset you or that either you would design a project to get only the answer you believe to be the best one or you might feel compelled to cover up findings that you don't like, then you need to choose a different topic. For example, one student wanted to find out whether there was any relationship between intelligence and political party affiliation. He was certain from the beginning that the members of his party were without a doubt the most intelligent. His strong opinion was not in and of itself the problem. However, his utter refusal to grant that it was even a possibility that the opposing party's members were more intelligent than those of his party, led him to decide that the topic was probably too near and dear for him to use it to conduct unbiased research.

Of course, just because you feel strongly about a topic does not mean that you should not study it. Sometimes the best topics to research are those about which you do feel strongly. What better way to stay motivated than to study something that you care about? I recently began a study of recreational user conflict in cottage country, precisely because I live in cottage country and it is an ever-present issue for the local population.

Although I have strong opinions about carrying-capacity, environmental degradation, and the interplay between locals, cottagers, and tourists, I also feel OK about having those ideas challenged. In fact, for me one of the most rewarding things about studying a topic that is relevant to my own life is learning new perspectives that had never occurred to me before collecting data on the topic. I believe that my own perspective is pretty solid, but I can also accept that other people will have perspectives that differ from my own. As well, I am certainly willing to report the variety of per-

spectives that I discover as I collect data on my topic and reach conclusions perhaps at odds with my initial thoughts. If you feel prepared to accept all findings, even those that may be unflattering to or distinct from your personal perspective, then perhaps you should intentionally study a topic about which you have strong feelings. However, if, after honest reflection, you decide that you cannot accept or share with others findings with which you disagree, then you should study a topic about which you feel less strongly.

What Do You Know About Where You Already Are?

Whether or not you feel strongly about your topic, you will also want to consider what you already know about it. There are many ways we know what we know. Perhaps your mother told you something is so. Perhaps it came to you in a dream. Perhaps you took a class last semester and learned something about your topic there. Or you may have read something about your topic in your local newspaper or a magazine. We discussed the strengths and weaknesses associated with some of these different sources of knowledge earlier, and we'll talk about other sources of knowledge, such as prior research, a little later on. For now, take some time to think about what you know about your topic from any and all possible sources. Thinking about what you already know will help you identify any biases you may have, and it will help as you begin to frame a question about your topic.

Is It Empirical?

When it comes to research questions, social scientists are best equipped to answer empirical questions—those that can be answered by real experience in the real world—as opposed to ethical questions—questions about which people have moral opinions and that may not be answerable in reference to the real world. While social scientists do study phenomena about which people have moral opinions, our job is to gather social facts about those phenomena, not to judge or determine morality.

Let's consider a specific example, health care. As you may be aware, different jurisdictions follow different models and pursue different overall aims regarding the funding and distribution of health care. Upon considering the various public, private, and mixed models of health care, one might be inclined to ask, "what is the most morally upstanding way to fund and allocate health care resources?"

Although this is an interesting and important question it is not, this question isn't empirical. But social scientists study inequality, one might argue, and understanding the most morally upright way of administering health care certainly had something to do with issues of inequality. This may indeed be true, but the problem was not with the topic per se but instead with the framing of the topic. As stated, it is an ethical question and not an empirical one. A related, empirical question may be "what type of system yields the best health-related outcomes" or something more specific like "do 5-year cancer survival rates vary by type of health system?" Although these are considerably smaller in scope, they have the advantage of being answerable through empirical research.

Of course, this doesn't mean that social scientists cannot study opinions about or social meanings. In fact, social scientists may be among the most qualified to gather empirical facts about people's moral opinions. We study humans after all, and as you will discover in the following chapters of this text, we are trained to utilize a variety of scientific data-collection techniques to understand patterns of human beliefs and behaviors. Using these techniques, we could find out how many people believe that abortion morally reprehensible, but we could never learn, empirically, whether abortion is, in fact, morally reprehensible.

Complementary Disciplines in Applied Human Sciences

It is important to consider that there a range of disciplinary perspectives that are complementary in the social sciences, each providing a particular lens through which to examine any given social phenomenon. In the applied human sciences, we are most familiar with theory and findings from sociology, psychology, and social-psychology, but disciplines such as economics, history, and political science can each contribute to our understanding.

For example, many students pursue double majors in sociology and psychology. While the two disciplines are complementary, they are not the same. Consider the topic of gang membership. While a psychologist may be interested in identifying what traumatic personal experiences or emotional state might drive a person to join a gang, a sociologist is more likely to examine whether there are patterns in terms of who joins gangs. Are members of some social classes more likely than others to join gangs? Does a person's geographical location appear to play a role in determining the likelihood that he or she will join a gang? In other words, psychologists and sociologists share an interest in human behavior, but psychologists tend to focus on individuals while sociologists consider individuals within the context of the social groups to which they belong.

Philosophers and social scientists also share some common interests, including a desire to understand beliefs about the nature of good and bad. But while a philosopher might consider what general or logical principles make up a good or a bad society, a social scientist is more likely to study how specific social realities, such as the presence of gangs in a community, impact perceptions of that community as either good or bad. Other disciplines that share some overlapping interests include political science, economics, and history.

Is It a Question?

Now that you've thought about what topics interest you and identified a topic that is both empirical and broadly concerned with the applied human sciences, you need to form a research question about that topic. [1] So what makes a good research question? First, it is generally written in the form of a question. To say that your research question is "child-free adults" or "students' knowledge about current events" or "movies" would not be correct. You need to frame a question about the topic that you wish to study. A good research question is also one that is well focused. Writing a well- focused question isn't really all that different from what the paparazzi do regularly. As a social scientist you need to be as clear and focused as those photographers who stalk Britney Spears to get that perfect shot of her while she waits in line at Starbucks. OK, maybe what we do as social scientists isn't exactly the same, but think about how the paparazzi get paid. They must take clear, focused photographs in order to get paid for what they do. Likewise, we will not hit the jackpot of having our research published, read, or respected by our peers if we are not clear and focused.

In addition to being written in the form of a question and being well focused, a good research question is one that cannot be answered with a simple yes or no. For example, if your interest is in gender norms, you could ask, "Does gender affect a person's shaving habits?" but you will have nothing left to say once you discover your yes or no answer. Instead, why not ask, "How or to what extent does gender affect a person's feelings about body hair?" By tweaking your question in this small way, you suddenly have a much more fascinating question and more to say as you attempt to answer it.

A good research question should also have more than one plausible answer. The student who studied the relationship between gender and body hair preferences had a specific interest in the impact of gender, but she also knew that preferences might vary on other dimensions. For example, she knew from her own experience that her more politically conservative friends were more likely to shave every day and more likely to only date other regular shavers. Thinking through the possible relationships between gender, politics, and shaving led that student to realize that there were many plausible answers to her questions about how gender affects a person's feelings about body hair. Because gender doesn't exist in a vacuum she, wisely, felt that she needed to take into account other characteristics that work together with gender to shape people's behaviors, likes, and dislikes. By doing this, the student took into account the third feature of a good research question: She considered relationships between several concepts. While she began with an interest in a single concept—body hair—by asking herself what other concepts (such as gender or political orientation) might be related to her original interest, she was able to form a question that considered the relationships among those concepts.

In sum, a good research question generally has the following features:

- It is written in the form of a question.

- It is clearly focused.
- It is not a yes/no question.
- It has more than one plausible answer.
- It considers relationships among multiple concepts.

In short, you will struggle unless you are clear in your aims and clearly focus your research question. You could be the most eloquent writer in your class, or even in the world, but if the research question about which you are writing is unclear, your work will ultimately fall flat. The good news is that much of this text is dedicated to learning how to write, and then answer, a good research question.

Some Specific Examples

Putting all this advice together, let's take a look at a few more examples of possible research questions and consider the relative strengths and weaknesses of each. While reading Table 3.1 below, keep in mind that I have only noted what I view to be the most relevant strengths and weaknesses of each question. Certainly, each question may have additional strengths and weaknesses not noted in the table. Also, it may interest you to know that the questions in Table 3.1 "Sample Sociological Research Questions: Strengths and Weaknesses" all come from undergraduate sociology student projects that I have either advised in the course of teaching sociological research methods or have become familiar with from sitting on undergraduate thesis committees. The work by thesis students is cited.

Table 3.1 Sample Research Questions: Strengths and Weaknesses

Sample Question	Question's strengths	Question's weaknesses	Proposed alternative
Do children's books teach us about gender norms in our society?	Written as a question. Focused.	Written as a yes/no.	What (or how) do children's books teach us about gender norms in our society?
Why are some men such jerks?	Written as a question. Focused.	Lacks theoretical grounding. Biased.	Who supports sexist attitudes and why?
Does sexual maturity change depending on where you're from?	Written as a question.	Unclear phrasing. Written as a yes/no.	How does knowledge about sex vary across different geographical regions?
What is sex?	Written as a question.	Too broadly focused. Not clear whether question is social scientific. Does not consider relationships among concepts.	How do students' definitions of sex change as they age?
Do social settings and peers and where you live influence an university student's exercise and eating habits?	Written as a question. Considers relationships among multiple concepts.	Lacks clarity. Unfocused. Written as a yes/no.	How does social setting influence a person's engagement in healthy behaviors?
What causes people to ignore someone in need of assistance?	Written as a question. Socially relevant.		
How do workers cope with short-term unemployment?	Written as a question. Focused. More than one plausible answer.		
Are motivations to volunteer gendered?	Written as a question. Socially relevant. More than one plausible answer.		
How have representations gender in video games changed over time?	Written as a question. Considers relationships among multiple concepts.		

Now that you have thought about topics that interest you and you've learned how to frame those topics empirically, as social science, and as questions, you have probably come up with a few potential research questions—questions to which you are dying to know the answers. However, even if you have identified the most brilliant research question ever, you are still not ready to begin conducting research. First, you'll need to think about the feasibility of your research question and to make a visit to your campus library.

Feasibility

We touched on the ethics of research earlier, but in addition to ethics there are a few practical matters related to feasibility that all researchers should consider before beginning a research project. Are you interested in better understanding the day-to-day experiences of maximum security prisoners? This sounds fascinating, but unless you plan to commit a crime that lands you in a maximum security prison, chances are good that you will not be able to gain access to this population. Perhaps your interest is in the inner workings of toddler peer groups. If you're much older than four or five, however, it might be tough for you to access even that sort of group. Your ideal research topic might require you to live on a chartered sailboat in the Bahamas for a few years, but unless you have unlimited funding, it will be difficult to make even that happen. The point, of course, is that while the topics about which questions can be asked may seem limitless, there are limits to which aspects of topics we can study, or at least to the ways we can study them.

In addition to your personal or demographic characteristics that could shape what you are able to study or how you are able to study it, there are also the very practical matters of time and money. In terms of time, your personal time-frame for conducting research may be the semester during which you are taking a class or a thesis year at grad school. Perhaps as an employee one day your employer will give you an even shorter timeline in which to conduct some research—or perhaps longer. How much time a researcher has to complete her or his work may depend on a number of factors and will certainly shape what sort of research that person is able to conduct. Money, as always, is also relevant. For example, your ability to conduct research while living on a chartered sailboat in the Bahamas may be hindered unless you have unlimited funds or win the lottery. And if you wish to conduct survey research, you may have to think about the fact that mailing paper surveys costs not only time but money—from printing them to paying for the postage required to mail them.

In sum, feasibility is always a factor when deciding what, where, when, and how to conduct research. Aspects of your own identity may play a role in determining what you can and cannot investigate, as will the availability of resources such as time and money.

How to Design a Research Project?

Now that you've figured out what to study, you need to figure out how to study it. Your library research can help in this regard. Reading published studies is a great way to familiarize yourself with the various components of a research project. It will also bring to your attention some of the major considerations to keep in mind when designing a research project. We'll say more about reviewing the literature later, but we'll begin with a focus on research design. We'll discuss the decisions you need to make about the goals of your research, the major components of a research project, along with a few additional aspects of designing research.

Goals of the Research Project

I have a 9-year-old daughter, whose grandparents recently bought her an iPad. As she has immersed herself with gusto, I have had any number of questions swirling around in my head. What sorts of gadgets are kids drawn to, or to what uses? How much is too much, and why? Do attitudes or behaviours of heavy users differ from those who are light users? How does a potential dependency develop, and who is most likely to experience one?

Social research is great for answering just these sorts of questions, but in order to answer our questions well, we must take care in designing our research projects. In this chapter, we'll consider what aspects of a research project should be considered at the beginning, including specifying the goals of the research, the components that are common across most research projects, and a few other considerations.

One of the first things to think about when designing a research project is what you hope to accomplish, in very general terms, by conducting the research. What do you hope to be able to say about your topic? Do you hope to gain a deep understanding of whatever phenomenon it is that you're studying, or would you rather have a broad, but perhaps less deep, understanding? Do you want your research to be used by policymakers or others to shape social life, or is this project more about exploring your curiosities? Your answers to each of these questions will shape your research design.

Exploration, Description, Explanation

You'll need to decide in the beginning phases whether your research will be exploratory, descriptive, or explanatory. Each has a different purpose, so how you design your research project will be determined in part by this decision.

Researchers conducting exploratory research are typically at the early stages of examining their topics. These sorts of projects are usually conducted when a researcher wants to test the feasibility of conducting a more extensive study; he or she wants to figure out the lay of the land, with respect to the particular topic. Perhaps very little prior research has been conducted on this subject. If this is the case, a researcher may wish to do some exploratory work to learn what method to use in collecting data, how best to approach research subjects, or even what sorts of questions are reasonable to ask. In the case of studying young people's dependency on their electronic gadgets, a researcher conducting exploratory research on this topic may simply wish to learn more about students' use of these gadgets. Because these dependencies seem to be a relatively new phenomenon, an exploratory study of the topic might make sense as an initial first step toward understanding it. As a further example, in my research on gentrification and adolescents, I was unsure what the results might be when first embarking on the study. There was very little empirical research on the topic, so the initial goal of the research was simply to explore how adolescents living in a gentrifying community perceived changes to their community, and how this may have affected their use of the community's resources. Conducting exploratory research on the topic was a necessary first step to understand the phenomenon better, in order to design a more tightly focused study subsequently.

Sometimes the goal of research is to describe or define a particular phenomenon. In this case, descriptive research would be an appropriate strategy. A descriptive study of university students' addictions to their electronic gadgets, for example, might aim to describe patterns in how use of gadgets varies by gender or university major or which sorts of gadgets students tend to use most regularly.

One example of descriptive research with which most of us are familiar are course evaluations. At least in pre- COVID days, students completed course evaluations at the end of each semester, and the consolidated results provided descriptive data about students' perceptions of assigned materials, the knowledge, abilities, and availability of their instructor, and their overall level of satisfaction with a course. These allow instructors to understand what they may be doing well, and what aspects may use improvement.

Finally, social science researchers often aim to explain why particular phenomena work in the way that they do. Research that answers "why" questions is referred to as explanatory research. In this case, the researcher is trying to identify the causes and effects of whatever phenomenon he or she is studying. An explanatory study of university students' addictions to their electronic gadgets

might aim to understand why students become addicted. Does it have anything to do with their family histories? With their other extracurricular hobbies and activities? With whom they spend their time? An explanatory study could answer these kinds of questions.

There are numerous examples of explanatory social scientific investigations. For example, Simons and Wurtele (2010) sought to discover whether receiving corporal punishment from parents led children to turn to violence in solving their interpersonal conflicts with other children. In their study of 102 families with children between the ages of 3 and 7, the researchers found that experiencing frequent spanking did, in fact, result in children being more likely to accept aggressive problem-solving techniques.

Another example of explanatory research can be seen in Faris and Felmlee's research (2011; American Sociological Association, 2011) on the connections between popularity and bullying. They found, from their study of 8th, 9th, and 10th graders in 19 North Carolina schools, that as adolescents' popularity increases, so, too, does their aggression.

Idiographic or Nomothetic?

Once you decide whether you will conduct exploratory, descriptive, or explanatory research, you will need to determine whether you want your research to be idiographic or nomothetic. A decision to conduct idiographic research means that you will attempt to explain or describe your phenomenon exhaustively. While you might have to sacrifice some breadth of understanding if you opt for an idiographic explanation, you will gain a deep, rich understanding of whatever phenomenon or group you are studying. In most cases, idiographic research falls to qualitative researchers, who seek far greater depth of understanding but who sacrifice generalizability due to the generally limited number of research participants. A decision to conduct nomothetic research, on the other hand, means that you will aim to provide a more general, sweeping explanation or description of your topic. In this case, you sacrifice depth of understanding in favor of breadth of understanding, while increasing the generality of the findings.

Applied or Basic?

Finally, you will need to decide what sort of contribution you hope to make with your research. Do you want others to be able to use your research to shape social life? If so, you may wish to conduct a study that policymakers could use to change or create a specific policy. Perhaps, on the other hand, you wish to conduct a study that will contribute to theories or knowledge without having a

specific applied use in mind. In the example of my daughter's burgeoning addiction to technological gadgets, an applied study of this topic might aim to understand how to treat such addictions. A basic study of the same topic, on the other hand, might examine existing theories of addiction and consider how this new type of addiction does or does not apply; perhaps your study could suggest ways that such theories may be tweaked to encompass technological addictions.

Earlier, we learned about both applied and basic research. When designing your research project, think about where you envision your work fitting in on the applied–basic continuum. Recognize, however, that even basic research may ultimately be used for some applied purpose. Similarly, your applied research might not turn out to be applicable to the particular real-world social problem you were trying to solve, but it might better our theoretical understanding of some phenomenon. In other words, deciding now whether your research will be basic or applied doesn't mean that will be its sole purpose forever. Basic research may ultimately be applied, and applied research can certainly contribute to general knowledge. Nevertheless, it is important to think in advance about what contribution(s) you hope to make with your research.

We have discussed the importance of understanding the differences between qualitative and quantitative research methods. Because this distinction is relevant to how researchers design their projects, we'll revisit it here.

Causality

Causality refers to the idea that one event, behavior, or belief will result in the occurrence of another, subsequent event, behavior, or belief. In other words, it is about cause and effect. In a quantitative study, a researcher is likely to aim for a nomothetic understanding of the phenomenon that he or she is investigating. In the case of media or device dependency, the researcher may be unable to identify the specific idiosyncrasies of individuals' particular patterns and perceptions of use. However, by analyzing data from a much larger and more representative group, the researcher will be able to identify the most likely, and more general, factors that account for students' addictions to electronic gadgets. The researcher might choose to collect survey data from a wide swath of university students from around the country. He might find that students who report addictive tendencies when it comes to their gadgets also tend to be people who participate across more social media platforms, are more likely to be men, and tend to engage in rude or disrespectful behaviors more often than peers without an addiction. It is possible, then, that these associations can be said to have some causal relationship to electronic gadget addiction. However, items that seem to be related are not necessarily causal. To be considered causally related in a nomothetic study, such as the survey research in this example, there are a few criteria that must be met.

The main criteria for causality have to do with ***plausibility, temporality, and spuriousness. Plausibility*** means that in order to make the claim that one event, behavior, or belief causes another, the claim has to make sense. For example, during a series of lectures, if certain students engage in midclass texting or web surfing even though they are aware this distracts others, one might begin to wonder whether people who are insensitive to others are more likely to exhibit dependency upon their electronic devices. However, the fact that there might be a relationship between insensitivity toward others and device dependence does not mean that a student's insensitivity could cause him to be device dependent. In other words, just because there might be some correlation between two variables does not mean that a causal relationship between the two is really plausible.

The criterion of ***temporality*** means that whatever cause you identify must precede its effect in time. For instance, a survey researcher examining the causes of students' digital device dependence might derive a number of findings. First, the researcher may find that those who identify as male exhibit greater device dependence than those who identify as female; that is, there is a relationship between gender identity and device dependence. In this case, one's gender identity is more likely to precede device dependence in time than device dependence is to precede one's gender identification. As a matter of logic, then, it may be able to establish the temporal order of the variables.

Alternately, consider the finding that the longer one has owned a smartphone, the more likely one is to exhibit device dependence. In this case, the researcher has found an association between duration of smartphone ownership and device dependence; however, what is the temporal order? It is equally plausible that those who are more device dependent will have contrived to get a smart-

phone earlier as it is to argue that the longer one has had a smartphone the more likely one will be device dependent. We will return to this point later when we discuss cross-sectional and experimental research.

Finally, a ***spurious*** relationship is one in which an association between two variables appears to be real but can in fact be explained by some third variable. Did you know, for example, that rates of ice cream sales have been shown to be related to the number of drowning deaths? Of course, it is not a true relationship. It is a mathematical artefact that arises because both drowning deaths and ice cream sales go up and down based on the level of a third variable. The third variable is time of year, across which both ice cream sales and drowning deaths rise or fall according to the temperature. Another classic example is that the more firefighters show up at a fire, the more damage is done at the scene. Of course, firefighters are not the cause of damage; rather, the amount of damage caused and the number of firefighters called on to help are both related to the size of the fire (Frankfort-Nachmias & Leon-Guerro, 2011). In each of these examples, it is the presence of a third variable that explains the apparent relationship between the two original variables.

In sum, the following criteria must be met in order for a correlation to be considered causal:

- The relationship must be plausible.
- The cause must precede the effect in time.
- The relationship must be nonspurious.

What we've been talking about here is relationships between variables. When one variable causes another, we have what researchers call independent and dependent variables. In the example where gender identity was found to be causally linked to electronic gadget addiction, gender would be the independent variable and electronic gadget addiction would be the dependent variable. An independent variable is one that causes another. A dependent variable is one that is caused by another. An easy was to remember this is that dependent variables depend on independent variables.

Relationship strength is another important factor to take into consideration when attempting to make causal claims if your research approach is nomothetic. I'm not talking strength of your friendships or marriage (though of course that sort of strength might affect your likelihood to keep your friends or stay married). In this context, relationship strength refers to statistical significance. The more statistically significant a relationship between two variables is shown to be, the greater confidence we can have in the strength of that relationship. We'll discuss statistical significance in greater detail in. For now, keep in mind that for a relationship to be considered causal, it cannot exist simply because of the chance selection of participants in a study.

Units of Analysis and Units of Observations

Another point to consider when designing a research project has to do with units of analysis and units of observation. These two items concern what you, the researcher, actually observe in the course of your data collection and what you hope to be able to say about those observations. A ***unit of analysis*** is the entity that you wish to be able to say something about at the end of your study, probably what you'd consider to be the main focus of your study. A ***unit of observation*** is the item (or items) that you actually observe, measure, or collect in the course of trying to learn something about your unit of analysis. In a given study, the unit of observation might be the same as the unit of analysis, but that is not always the case. Further, units of analysis are not required to be the same as units of observation. What is required, however, is for researchers to be clear about how they define their units of analysis and observation, both to themselves and to their audiences.

More specifically, your unit of analysis will be determined by your research question. Your unit of observation, on the other hand, is determined largely by the method of data collection that you use to answer that research question. We'll take a closer look at methods of data collection in later chapters. For now, let's go back to the example we've been discussing over the course of this chapter, students' electronic device dependence. We'll consider first how different kinds of research questions about this topic will yield different units of analysis. Then we'll think about how those questions might be answered and with what kinds of data. This leads us to a variety of units of observation.

If we were to ask, "Which students are most likely to exhibit dependence on their digital device?" our unit of observation would be the individual. We might mail a survey to students on campus, and our aim might be to determine whether membership in certain programs of study might be related to device dependence. We might find that majors in Communication Studies, Computational Arts, and Software Engineering are all more likely than other students to become dependent on their digital devices. As you will note, the unit of analysis is "program of study." Although each program is made up of students, for the purposes of analysis, we are interested in the group, not the individuals.

Indeed, a common unit of analysis in social scientific inquiry is groups. Groups of course vary in size, and almost no group is too small or too large to be of interest. Families, friendship groups, and street gangs make up some of the more common micro-level groups examined by social scientists. Employees in an organization, professionals in a particular domain (e.g., chefs, lawyers, social scientists), and members of clubs (e.g., Girl Scouts, Rotary, Red Hat Society) are all meso-level groups that may function as units of analyses. Finally, at the macro level, it is possible to examine citizens of entire nations or residents of different continents or other regions.

At the group level, a study of student dependence on their smart devices might consider whether certain types of social clubs have more or fewer gadget-addicted members than other sorts of clubs. Perhaps we would find that clubs that emphasize physical fitness, such as the rugby

club and the scuba club, have fewer gadget-addicted members than clubs that emphasize cerebral activity, such as the chess club and the sociology club. Our unit of analysis in this example is groups. If we had instead asked whether people who join cerebral clubs are more likely to be gadget-addicted than those who join social clubs, then our unit of analysis would have been individuals. In either case, however, our unit of observation would be individuals.

Organizations are yet another potential unit of analysis that social scientists might wish to say something about. Organizations include entities like corporations, universitys and universities, and even night clubs. At the organization level, a study of students' device dependence might ask, "How do different universitys address the problem of device dependence?" In this case, our interest lies not in the experience of individual students but instead in the campus- to-campus differences in confronting device dependence. A researcher conducting a study of this type might examine schools' written policies and procedures, so his unit of observation might be documents, key administrative personnel, or service providers. However, because he ultimately wishes to describe differences across universities, the university would be his unit of analysis.

Of course, it would be silly in a textbook focused on social scientific research to neglect social phenomena as a potential unit of analysis. I mentioned one such example earlier, but let's look more closely at this sort of unit of analysis. Many social scientists study a variety of social interactions and social problems that fall under this category. Examples include social problems like murder or rape; interactions such as counseling sessions, Facebook chatting, or wrestling; and other social phenomena such as voting and even gadget use or misuse. A researcher interested in students' electronic device dependence could ask, "What are the various types of device dependence that exist among students?" Perhaps the researcher will discover that some dependencies are primarily centered around social media or texting while other dependencies center more on various iterations of gaming. The resultant typology of device dependencies would tell us something about the social phenomenon (unit of analysis) being studied. As in several of the preceding examples, however, the unit of observation would likely be individual people.

Finally, a number of social scientists examine policies and principles, the last type of unit of analysis we'll consider here. Studies that analyze policies and principles typically rely on documents as the unit of observation. Perhaps a researcher has been hired by a university to help it write an effective policy to guard against device dependence. In this case, the researcher might gather all previously written policies from campuses all over the country and compare policies at campuses where device dependence rates are low to policies at campuses where device dependence rates are high. In sum, there are many potential units of analysis that a social scientist might examine, but some of the most common units include the following:

- Individuals
- Groups
- Organizations
- Social phenomena
- Policies and principles

Errors in Logical Reasoning

A ***fallacy*** is an error in logical reasoning. This is to say, the conclusion that one has drawn is not logically supported by it premises. There are a great number of logical fallacies, but there are a couple that are of specific relevance to this discussion. One common error we see people make when it comes to both causality and units of analysis is something called the ***ecological fallacy***. This occurs when claims about one lower-level unit of analysis are made based on data from some higher-level unit of analysis. In many cases, this occurs when claims are made about individuals, but only group-level data have been gathered. For example, we might want to understand whether electronic gadget addictions are more common on certain campuses than on others. Perhaps different campuses around the country have provided us with their campus percentage of gadget-addicted students, and we learn from these data that electronic gadget addictions are more common on campuses that have business programs than on campuses without them. We then conclude that business students are more likely than nonbusiness students to become addicted to their electronic gadgets. However, this would be an inappropriate conclusion to draw. Because we only have addiction rates by campus, we can only draw conclusions about campuses, not about the individual students on those campuses. Perhaps the sociology majors on the business campuses are the ones that caused the addiction rates on those campuses to be so high. The point is we simply don't know because we only have campus- level data. By drawing conclusions about students when our data are about campuses, we run the risk of committing the ecological fallacy.

On the other hand, another mistake to be aware of is ***reductionism***. Reductionism occurs when claims about some higher-level unit of analysis are made based on data from some lower-level unit of analysis. In this case, claims about groups or macro-level phenomena are made based on individual-level data. An example of reductionism can be seen in some descriptions of the start of World War 1. Many have opined that the WW1 was caused by the assassination of Archduke Ferdinand. Although it is true that this was a catalyzing event, it is reductionist to proclaim it as the cause of WW1. Obviously, geopolitical events on the grand scale need to align just right to initiate a global conflict. Did the assassination in and of itself did she cause the war? To say yes would be reductionist.

It would be a mistake to conclude from the preceding discussion that researchers should avoid making any claims whatsoever about data or about relationships between variables. While it is important to be attentive to the possibility for error in causal reasoning about different levels of analysis, this warning should not prevent you from drawing well-reasoned analytic conclusions from your data. The point is to be cautious but not abandon entirely the social scientific quest to understand patterns of behavior.

Hypothesis

In some cases, the purpose of research is to test a specific hypothesis or hypotheses. At other times, researchers do not have predictions about what they will find but instead conduct research to answer a question or questions, with an open-minded desire to know about a topic, or to help develop hypotheses for later testing.

An ***hypothesis*** is a statement, sometimes but not always causal, describing a researcher's expectation regarding what he or she anticipates finding. Often hypotheses are written to describe the expected relationship between two variables (though this is not a requirement). To develop a hypothesis, one needs to have an understanding of the differences between independent and dependent variables and between units of observation and units of analysis. Hypotheses are typically drawn from theories and usually describe how an independent variable is expected to affect some dependent variable or variables. Researchers following a deductive approach to their research will hypothesize about what they expect to find based on the theory or theories that frame their study. If the theory accurately reflects the phenomenon it is designed to explain, then the researcher's hypotheses about what he or she will observe in the real world should bear out.

Let's consider a couple of examples. Based on feminist theories of sexual harassment, one may hypothesize that "more females than males will experience specific sexually harassing behaviors." What is the causal relationship being predicted here? Which is the independent and which is the dependent variable? In this case, we hypothesized that a person's sex (independent variable) would predict her or his likelihood to experience sexual harassment (dependent variable).

Sometimes researchers will hypothesize that a relationship will take a specific direction. As a result, an increase or decrease in one area might be said to cause an increase or decrease in another. For example, you might choose to study the relationship between age and legalization of marijuana. Perhaps you've done some reading in this area and, based on the theories you've read, you hypothesize that "age is negatively related to support for marijuana legalization." What have you just hypothesized? You have hypothesized that as people get older, the likelihood of their supporting marijuana legalization decreases. Thus as age (your independent variable) moves in one direction (up), support for marijuana legalization (your dependent variable) moves in another direction (down).

Note that even with the most compelling data, you will almost never hear researchers say that they have proven their hypotheses. A statement that bold implies that a relationship has been shown to exist with absolute certainty and that there is no chance that there are conditions under which the hypothesis would not bear out. Instead, researchers tend to say that their hypotheses have been supported (or not). This more cautious way of discussing findings allows for the possibility that new evidence or new ways of examining a relationship will be discovered.

Researchers may also discuss a null hypothesis, one that predicts no relationship between the variables being studied. If a researcher rejects the null hypothesis, he or she is saying that the variables in question are somehow related to one another. We will have more to say about this when we discuss hypothesis testing.

Key Takeaways, Exercises, and References

Key Takeaways

- Many researchers choose topics by considering their own personal experiences, knowledge, and interests.
- Researchers should be aware of and forthcoming about any strong feelings they might have about their research topics.
- There are benefits and drawbacks associated with studying a topic about which you already have some prior knowledge or experience. Researchers should be aware of and consider both.
- Empirical questions are distinct from ethical questions.
- There are usually a number of ethical questions and a number of empirical questions that could be asked about any single topic.
- While social scientists may study topics about which people have moral opinions, their job is to gather empirical data about the social world.
- When thinking about the feasibility of their research questions, researchers should consider their own identities and characteristics along with any potential constraints related to time and money.
- Most strong research questions have five key features: written in the form of a question, clearly focused, beyond yes/no, more than one plausible answer, and consider relationships among concepts.
- A poorly focused research question can lead to the demise of an otherwise well-executed study.
- In qualitative studies, the goal is generally to understand the multitude of causes that account for the specific instance the researcher is investigating.
- In quantitative studies, the goal may be to understand the more general causes of some phenomenon rather than the idiosyncrasies of one particular instance.
- In order for a relationship to be considered causal, it must be plausible and nonspurious, and the cause must precede the effect in time.
- A unit of analysis is the item you wish to be able to say something about at the end of your study while a unit of observation is the item that you actually observe.
- When researchers confuse their units of analysis and observation, they may be prone to committing either the ecological fallacy or reductionism.
- Hypotheses are statements, drawn from theory, which describe a researcher's expectation about a relationship between two or more variables.
- Exploratory research is usually conducted when a researcher has just begun an investigation

and wishes to understand her or his topic generally.
- Descriptive research is research that aims to describe or define the topic at hand.
- Explanatory research is research that aims to explain why particular phenomena work in the way that they do.
- Idiographic investigations are exhaustive; nomothetic investigations are more general.
- Applied research may contribute to basic understandings and that basic research may also turn out to have some useful application.
- Applied Human Sciences appeal to a variety of disciplinary perspectives
- Though different disciplines address similar topics, there are distinct features that separate each discipline

Exercises

- Do some brainstorming to try to identify some potential topics of interest. What have been your favorite classes in university thus far? What did you like about them? What did you learn in them? What extracurricular activities are you involved in? How do you enjoy spending your time when nobody is telling you what you should be doing?
- Pick two variables that are of interest to you (e.g., age and religiosity, gender and university major, geographical location and preferred sports). State a hypothesis that specifies what you expect the relationship between those two variables to be. Now draw your hypothesis.
- Name a topic that interests you. Now keeping the features of a good research question in mind, come up with three possible research questions you could ask about that topic.
- Discuss your topic with a friend or with a peer in your class. Ask that person what sorts of questions come to mind when he or she thinks about the topic. Also ask that person for advice on how you might better focus one or all the possible research questions you came up with on your own.
- Think of some topic that interests you. Pose one ethical question about that topic. Now pose an empirical question about the same topic.
- Read a few news articles about any controversial topic that interests you (e.g., immigration, gay marriage, health care reform, terrorism, welfare). Make a note of the ethical points or questions that are raised in the articles and compare them to the empirical points or questions that are mentioned. Which do you find most compelling? Why?
- Describe a scenario in which exploratory research might be the best approach. Similarly, describe a scenario in which descriptive and then explanatory research would be the preferred approach.
- Which are you more drawn to personally, applied or basic research? Why?
- Take a look around you the next time you are heading across campus or waiting in line at the

grocery store or your favorite coffee shop. Think about how the very experience you are having in that moment may be different for those around you who are not like you. How would a change in your physical capabilities alter your path across campus? Would you interpret the stares from the child sitting in her parents' cart at the grocery store differently if you were a different race?

References

Babbie, E. (2010). The practice of social research (12th ed.). Belmont, CA: Wadsworth.

Blee, K. (1991). Women of the klan: Racism and gender in the 1920s. Berkeley, CA: University of California Press.

Blee, K. (2002). Inside organized racism: Women and men of the hate movement. Berkeley, CA: University of California Press;

Esterberg, K. G. (2002). Qualitative methods in social research. Boston, MA: McGraw-Hill;

Frankfort-Nachmias, C., & Leon-Guerro, A. (2011). Social statistics for a diverse society (6th ed.). Thousand Oaks, CA: Pine Forge Press.

Huff, D., & Geis, I. (1993). How to lie with statistics. New York, NY: Norton.

Lofland, J., & Lofland, L. H. (1995). Analyzing social settings: A guide to qualitative observation and analysis (3rd ed.). Belmont, CA: Wadsworth.

MacLeod, J. (2008). Ain't no makin' it: Aspirations and attainment in a low-income neighborhood (3rd ed.). Boulder, CO: Westview Press.

CHAPTER 4: CONDUCTING A LITERATURE REVIEW

Whether you plan to engage in clinical, community, or organizational practice, being able to look at the available literature on a topic and synthesize the relevant facts into a coherent review is a fundamental skill. Literature reviews can have a powerful effect, for example by providing the factual basis for a new program or policy in an agency or government. In your own research proposal, conducting a thorough literature review will help you build strong arguments for why your topic is important and why your research question must be answered.

LEARNING OBJECTIVES

- Be able to describe a literature review and explain its purpose
- Become familiar with the notion of synthesizing literature and means to accomplish this
- Be aware of the beneficial practices and common pitfalls while writing the literature review

Chapter 4: Conducting a Literature Review is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

What is a Literature Review?

Pick up nearly any book on research methods and you will find a description of a literature review. At a basic level, the term implies a survey of factual or nonfiction books, articles, and other documents published on a particular subject. Definitions may be similar across the disciplines, with new types and definitions continuing to emerge. Generally speaking, a literature review is a:

"comprehensive background of the literature within the interested topic area" (O'Gorman & MacIntosh, 2015, p. 31).

"critical component of the research process that provides an in-depth analysis of recently published research findings in specifically identified areas of interest" (Houser, 2018, p. 109).

"written document that presents a logically argued case founded on a comprehensive understanding of the current state of knowledge about a topic of study" (Machi & McEvoy, 2012, p. 4).

Literature reviews are indispensable for academic research. "A substantive, thorough, sophisticated literature review is a precondition for doing substantive, thorough, sophisticated research…A researcher cannot perform significant research without first understanding the literature in the field" (Boote & Beile, 2005, p. 3). In the literature review, a researcher shows she is familiar with a body of knowledge and thereby establishes her credibility with a reader. The literature review shows how previous research is linked to the author's project, summarizing and synthesizing what is known while identifying gaps in the knowledge base, facilitating theory development, closing areas where enough research already exists, and uncovering areas where more research is needed. (Webster & Watson, 2002, p. xiii). They are often necessary for real world social work practice. Grant proposals, advocacy briefs, and evidence-based practice rely on a review of the literature to accomplish practice goals.

A literature review is a compilation of the most significant previously published research on your topic. Unlike an annotated bibliography or a research paper you may have written in other classes, your literature review will outline, evaluate, and synthesize relevant research and relate those sources to your own research question. It is much more than a summary of all the related literature. A good literature review lays the foundation for the importance of the problem your research project addresses defines the main ideas in your research question and their interrelationships.

Literature Review Basics

All literature reviews will at some point:

- Introduce the topic and define its key terms.
- Establish the importance of the topic.
- Provide an overview of the important literature on the concepts in the research question and other related concepts.
- Identify gaps in the literature or controversies.
- Point out consistent finding across studies.
- Arrive at a synthesis that organizes what is known about a topic, rather than just summarizing.
- Discusses possible implications and directions for future research.

There are many different types of literature reviews, including those that focus solely on methodology, those that are more conceptual, and those that are more exploratory. Regardless of the type of literature review or how many sources it contains, strong literature reviews have similar characteristics. Your literature review is, at its most fundamental level, an original work based on an extensive critical examination and synthesis of the relevant literature on a topic. As a study of the research on a particular topic, it is arranged by key themes or findings, which should lead up to or link to the research question.

A literature review is a mandatory part of any research project. It demonstrates that you can systematically explore the research in your topic area, read and analyze the literature on the topic, use it to inform your own work, and gather enough knowledge about the topic to conduct a research project. Literature reviews should be reasonably complete, and not restricted to a few journals, a few years, or a specific methodology or research design. A well-conducted literature review should indicate to you whether your initial research questions have already been addressed in the literature, whether there are newer or more interesting research questions available, and whether the original research questions should be modified or changed in light of findings of the literature review. The review can also provide some intuitions or potential answers to the questions of interest and/or help identify theories that have previously been used to address similar questions and may provide evidence to inform policy or decision-making (Bhattacherjee, 2012).

Literature reviews are also beneficial to you as a researcher and scholar in professional practice. By reading what others have argued and found in their work, you become familiar with how people talk about and understand your topic. You will also refine your writing skills and your understanding of the topic you have chosen. The literature review also impacts the question you want to answer. As you learn more about your topic, you will clarify and redefine the research question guiding your inquiry. Literature reviews make sure you are not "reinventing the wheel" by repeating a study done so many times before or making an obvious error that others have encountered. The contribution your research study will have depends on what others have found before you. Try to place the study you wish to do in the context of previous research and ask, "Is this contributing something new?" and "Am I addressing a gap in knowledge or controversy in the literature?"

In summary, you should conduct a literature review to:

- Locate gaps in the literature of your discipline

- Avoid “reinventing the wheel”
- Carry on the unfinished work of other scholars
- Identify other people working in the same field
- Increase breadth and depth of knowledge in your subject area
- Read the seminal works in your field
- Provide intellectual context for your own work
- Acknowledge opposing viewpoints
- Put your work in perspective
- Demonstrate you can find and understand previous work in the area

Common Literature Review Errors

Literature reviews are more than a summary of the publications you find on a topic. As you have seen in this brief introduction, literature reviews are a very specific type of research, analysis, and writing. We will explore these topics more in the next chapters. As you begin your literature review, here are some common errors to avoid:

- Accepting another researcher’s finding as valid without evaluating methodology and data
- Ignoring contrary findings and alternative interpretations
- Using findings that are not clearly related to your own study or using findings that are too general
- Dedicating insufficient time to literature searching
- Simply reporting isolated statistical results, rather than synthesizing the results
- Relying too heavily on secondary sources
- Overusing quotations from sources
- Not justifying arguments using specific facts or theories from the literature

Chapter 4: Conducting a Literature Review is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

Synthesizing Literature: Putting the Pieces Together

Combining separate elements into a whole is the dictionary definition of synthesis. It is a way to make connections among and between numerous and varied source materials. A literature review is not an annotated bibliography, organized by title, author, or date of publication. Rather, it is grouped by topic and argument to create a whole view of the literature relevant to your research question.

Your synthesis must demonstrate a critical analysis of the papers you collected, as well as your ability to integrate the results of your analysis into your own literature review. Each source you collect should be critically evaluated for relevance and quality before you include it in your review.

Begin the synthesis process by creating a grid, table, or an outline where you will summarize your literature review findings, using common themes you have identified and the sources you have found. The summary, grid, or outline will help you compare and contrast the themes, so you can see the relationships among them as well as areas where you may need to do more searching. Whichever method you choose, this type of organization will help you to both understand the information you find and structure the writing of your review. Remember, although "the means of summarizing can vary, the key at this point is to make sure you understand what you've found and how it relates to your topic and research question" (Bennard et al., 2014, para. 10).

As you read through the material you gather, look for common themes as they may provide the structure for your literature review. Remember, research is an iterative process. It is not unusual to go back and search academic databases for more sources of information as you read the articles you've collected.

Literature reviews can be organized sequentially or by topic, theme, method, results, theory, or argument. It's important to develop categories that are meaningful and relevant to your research question. Take detailed notes on each article and use a consistent format for capturing all the information each article provides. These notes and the summary table can be done manually using note cards. However, given the amount of information you will be recording, an electronic file created in a word processing or spreadsheet is more manageable. Examples of fields you may want to capture in your notes include:

- Authors' names
- Article title
- Publication year
- Main purpose of the article
- Methodology or research design
- Participants

- Variables
- Measurement
- Results
- Conclusions

Other fields that will be useful when you begin to synthesize the sum total of your research:

- Specific details of the article or research that are especially relevant to your study
- Key terms and definitions
- Statistics
- Strengths or weaknesses in research design
- Relationships to other studies
- Possible gaps in the research or literature (for example, many research articles conclude with the statement "more research is needed in this area")
- Finally, note how closely each article relates to your topic. You may want to rank these as high, medium, or low relevance. For papers that you decide not to include, you may want to note your reasoning for exclusion, such as small sample size, local case study, or lacks evidence to support conclusions.

An example of how to organize summary tables by author or theme is shown in Table 4.1.

Table 4.1: Summary Table

Author/ Year	Research Design	Participants or Population Studied	Comparison	Outcome
Smith/2010	Mixed methods	Undergraduates	Graduates	Improved access
King/2016	Survey	Females	Males	Increased representation
Miller/2011	Content analysis	Nurses	Doctors	New procedure

Creating a topical outline

An alternative way to organize your articles for synthesis it to create an outline. After you have collected the articles you intend to use (and have put aside the ones you won't be using), it's time to extract as much as possible from the facts provided in those articles. You are starting your research project without a lot of hard facts on the topics you want to study, and by using the literature reviews provided in academic journal articles, you can gain a lot of knowledge about a topic in a short period of time.

As you read an article in detail, I suggest copying the information you find relevant to your research topic in a separate word processing document. Copying and pasting from PDF to Word can be a pain because PDFs are image files not documents. To make that easier, use the HTML version of the article, convert the PDF to Word in Adobe Acrobat or another PDF reader, or use "paste special" command to paste the content into Word without formatting. If it's an old PDF, you may have to simply type out the information you need. It can be a messy job, but having all of your facts in one place is very helpful for drafting your literature review.

You should copy and paste any fact or argument you consider important. Some good examples include definitions of concepts, statistics about the size of the social problem, and empirical evidence about the key variables in the research question, among countless others. It's a good idea to consult with your professor and the syllabus for the course about what they are looking for when they read your literature review. Facts for your literature review are principally found in the introduction, results, and discussion section of an empirical article or at any point in a non- empirical article. Copy and paste into your notes anything you may want to use in your literature review.

Importantly, *you must make sure you note the original source of that information*! Nothing is worse than searching your articles for hours only to realize you forgot to note where your facts came from. If you found a statistic that the author used in the introduction, it almost certainly came from another source that the author cited in a footnote or internal citation. You will want to check the original source to make sure the author represented the information correctly. Moreover, you may want to read the original study to learn more about your topic and discover other sources relevant to your inquiry.

Assuming you have pulled all of the facts out of multiple articles, it's time to start thinking about how these pieces of information relate to each other. Start grouping each fact into categories and subcategories. For example, a statistic stating that homeless single adults are more likely to be male may fit into a category of gender and homelessness. For each topic or subtopic you identified during your critical analysis of each paper, determine what those papers have in common. Likewise, determine which ones in the group differ. If there are contradictory findings, you may be able to identify methodological or theoretical differences that could account for the contradiction. For

example, one study may sample only high-income earners or those in a rural area. Determine what general conclusions you can report about the topic or subtopic, based on all of the information you've found.

Create a separate document containing a topical outline that combines your facts from each source and organizes them by topic or category. As you include more facts and more sources into your topical outline, you will begin to see how each fact fits into a category and how categories relate to each other. Your category names may change over time, as may their definitions. This is a natural reflection of your learning. A complete topical outline is a long list of facts, arranged by category about your topic. As you step back from the outline, you should understand the topic areas where you have enough information to make strong conclusions about what the literature says. You should also assess in what areas you need to do more research before you can write a robust literature review. The topical outline should serve as a transitional document between the notes you write on each source and the literature review you submit to your professor. It is important to note that they contain plagiarized information that is copied and pasted directly from the primary sources. That's okay because these are just notes and are not meant to be turned in as your own ideas. For your final literature review, you must paraphrase these sources to avoid plagiarism. More importantly, you should keep your voice and ideas front-and-center in what you write as this is your analysis of the literature. Make strong claims and support them thoroughly using facts you found in the literature. We will pick up the task of writing your literature review later.

Additional resources for synthesizing literature

There are many ways to approach synthesizing literature. We've reviewed two examples here: summary tables and topical outlines. Other examples you may encounter include annotated bibliographies and synthesis matrixes. As you are learning research, find a method that works for you. Reviewing the literature is a core component of evidence-based practice in social work at any level. See the resources below if you need some additional help:

- https://library.concordia.ca/help/writing/literature-review.php
- Further resources are listed here
- Killam, Laura (2013). Literature review preparation: Creating a summary table. Includes transcript. https://www.youtube.com/watch?v=nX2R9FzYhT0

Chapter 4: Conducting a Literature Review is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

Writing the Literature Review

By now, you should have discovered, retrieved, evaluated, synthesized, and organized the information you need for your literature review. It's now time to turn that stack of articles, papers, and notes into a literature review–it's time to start writing! If you've followed the steps in this chapter, you likely have an outline from which you can begin the writing process. But what do you need to include in your literature review? We've mentioned it before here, but just to summarize, a literature review should:

...clearly describe the questions that are being asked. They also locate the research within the ongoing scholarly dialogue. This is done by summarizing current understandings and by discussing why what we already knows leads to the need for the present research. Literature reviews also define the primary concepts. While this information can appear in any order, these are the elements in all literature reviews. (Loseke, 2017, p. 61)

Do you have enough facts and sources to accomplish these tasks? It's a good time to consult your outlines and notes on each article you plan to include in your literature review. You may also want to consult with your professor on what they expect from you. If there is something that you are missing, you may want to jump back to section 2.3 where we discussed how to search for literature on your topic. While you can always fill in material later, there is always the danger that you will start writing without really knowing what you are talking about or what you want to say. For example, if you don't have a solid definition of your key concepts or a sense of how the literature has developed over time, it will be difficult to make coherent scholarly claims about your topic.

There is no magical point at which everyone is ready to write. As you consider whether you are ready or not, it may be useful to ask yourself these questions:

- How will my literature review be organized?
- What section headings will I be using?
- How do the various studies relate to each other?
- What contributions do they make to the field?
- What are the limitations of a study/where are the gaps in the research?
- And finally, but most importantly, how does my own research fit into what has already been done?

The problem statement

Many scholarly works begin with a problem statement. The problem statement serves two functions. On one hand, it establishes why your topic is a social problem worth studying. At the same time, it also pulls your reader into the literature review. Who would want to read about something unimportant?

A problem statement generally answers the following questions, though these are far from exhaustive:

- Why is this an important problem to study?
- How many people are affected by the problem?
- How does this problem impact other social issues or target populations relevant to social work?
- Why is your target population an important one to study?

A strong problem statement, like the rest of your literature review, should be filled with facts, theory, and arguments based on the literature you've found. A research proposal differs significantly from other more reflective essays you've likely completed during your social work studies. If your topic were domestic violence in rural Appalachia in the USA, I'm sure you could come up with answers to the above questions without looking at a single source. However, the purpose of the literature review is not to test your intuition, personal experience, or empathy. Instead, research methods are about learning specific and articulable facts to inform action. With a problem statement, you can take a "boring" topic like the color of rooms used in an inpatient psychiatric facility, transportation patterns in major cities, or the materials used to manufacture baby bottles and help others see the topic as you see it—an important part of the social world that impacts people's lived experience.

The structure of a literature review

The problem statement generally belongs at the beginning of the literature review. Take care not to go on for too long. I usually advise my students to spend no more than a paragraph or two for a problem statement. For the rest of your literature review, there is no set formula for how it should be organized. However, a literature review generally follows the format of any other essay—Introduction, Body, and Conclusion.

The introduction to the literature review contains a statement or statements about the overall topic. At minimum, the introduction should define or identify the general topic, issue, or area of concern. You might consider presenting historical background, mention the results of a seminal study, and provide definitions of important terms. The introduction may also point to overall trends in what has been previously published on the topic or conflicts in theory, methodology, evidence, conclusions, or gaps in research and scholarship. I also suggest putting in a few sentences that walk the reader through the rest of the literature review. Highlight your main arguments from the body of the literature review and preview your conclusion. An introduction should let someone know what to expect from the rest of your review.

The body of your literature review is where you demonstrate your synthesis and analysis of the literature on your topic. Again, take care not to just summarize your literature. I would also caution against organizing your literature review by source—that is, one paragraph for source A, one paragraph for source B, etc. That structure will likely provide an okay summary of the literature you've found, but it would give you almost no synthesis of the literature. That approach doesn't tell your reader how to put those facts together, points of agreement or contention in the literature, or how each study builds on the work of others. In short, it does not demonstrate critical thinking.

Instead, use your outlines and notes as a guide to the important topics you need to cover, and more importantly, what you have to say about those topics. Literature reviews are written from the perspective of an expert on the field. After an exhaustive literature review, you should feel like you are able to make strong claims about what is true—so make them! There is no need to hide behind "I believe" or "I think." Put your voice out in front, loud and proud! But make sure you have facts and sources that back up your claims.

I've used the term "argument" here in a specific way. An argument in writing means more than simply disagreeing with what someone else said. Toulman, Rieke, and Janik (1984) identify six elements of an argument:

- Claim: the thesis statement—what you are trying to prove
- Grounds: theoretical or empirical evidence that supports your claim
- Warrant: your reasoning (rule or principle) connecting the claim and its grounds
- Backing: further facts used to support or legitimize the warrant
- Qualifier: acknowledging that the argument may not be true for all cases
- Rebuttal: considering both sides (as cited in Burnette, 2012) 2

Let's walk through an example of an argument. If I were writing a literature review on a negative income tax, a policy in which people in poverty receive an unconditional cash stipend from the government each month equal to the federal poverty level. I would want to lay out the following:

- Claim: the negative income tax is superior to other forms of anti-poverty assistance.
- Grounds: data comparing negative income tax recipients to those in existing programs, theory supporting a negative income tax, data from evaluations of existing anti-poverty programs,

etc.

- Warrant: cash-based programs like the negative income tax are superior to existing anti-poverty programs because they allow the recipient greater self-determination over how to spend their money.
- Backing: data demonstrating the beneficial effects of self-determination on people in poverty.
- Qualifier: the negative income tax does not provide taxpayers and voters with enough control to make sure people in poverty are not wasting financial assistance on frivolous items.
- Rebuttal: policy should be about empowering the oppressed, not protecting the taxpayer, and there are ways of addressing taxpayer opposition through policy design.

Like any effective argument, your literature review must have some kind of structure. For example, it might begin by describing a phenomenon in a general way along with several studies that provide some detail, then describing two or more competing theories of the phenomenon, and finally presenting a hypothesis to test one or more of the theories. Or, it might describe one phenomenon, then describe another phenomenon that seems inconsistent with the first one, then propose a theory that resolves the inconsistency, and finally present a hypothesis to test that theory. In applied research, it might describe a phenomenon or theory, then describe how that phenomenon or theory applies to some important real-world situation, and finally suggest a way to test whether it does, in fact, apply to that situation.

Another important issue is ***signposting***. It may not be a term you are familiar with, but you are likely familiar with the concept. Signposting refers to the words used to identify the organization and structure of your literature review to your reader. The most basic form of signposting is using a topic sentence at the beginning of each paragraph. A topic sentence introduces the argument you plan to make in that paragraph. For example, you might start a paragraph stating, "There is strong disagreement in the literature as to whether psychedelic drugs cause psychotic disorders, or whether psychotic disorders cause people to use psychedelic drugs." Within that paragraph, your reader would likely assume you will present evidence for both arguments. The concluding sentence of your paragraph should address the topic sentence, addressing how the facts and arguments from other authors support a specific conclusion. To continue with our example, I might say, "There is likely a reciprocal effect in which both the use of psychedelic drugs worsens pre- psychotic symptoms and worsening psychosis causes use of psychedelic drugs to self-medicate or escape."

Signposting also involves using headings and subheadings. Your literature review will use APA formatting, which means you need to follow their rules for bolding, capitalization, italicization, and indentation of headings. Headings help your reader understand the structure of your literature review. They can also help if the reader gets lost and needs to re- orient themselves within the document. I often tell my students to assume I know nothing (they don't mind) and need to be shown exactly where they are addressing each part of the literature review. It's like walking a small child around, telling them "First we'll do this, then we'll do that, and when we're done, we'll know this!"

Another way to use signposting is to open each paragraph with a sentence that links the topic of the paragraph with the one before it. Alternatively, one could end each paragraph with a sentence that links it with the next paragraph. For example, imagine we wanted to link a paragraph about barriers to accessing healthcare with one about the relationship between the patient and physician. We could use a transition sentence like this: "Even if patients overcome these barriers to accessing care, the physician-patient relationship can create new barriers to positive health outcomes." A transition sentence like this builds a connection between two distinct topics. Transition sentences are also useful within paragraphs. They tell the reader how to consider one piece of information in light of previous information. Even simple transitions like however, similarly, and others demonstrate critical thinking and make your arguments clearer.

Many beginning researchers have difficulty with incorporating transitions into their writing. Let's look at an example. Instead of beginning a sentence or paragraph by launching into a description of a study, such as "Williams (2004) found that…," it is better to start by indicating something about why you are describing this particular study. Here are some simple examples:

- Another example of this phenomenon comes from the work of Williams (2004).
- Williams (2004) offers one explanation of this phenomenon.
- An alternative perspective has been provided by Williams (2004).

Now that we know to use signposts, the natural question is "What goes on the signposts?" First, it is extremely important to start with an outline of the main points that you want to make, organized in the order that you want to make them. The basic structure of your argument then should be apparent from the outline itself. Unfortunately, there is no formula I can give you that will work for everyone. I can provide some general pointers on structuring your literature review, though.

The literature review generally moves from general ideas to more specific ones. You can build a review by identifying areas of consensus and areas of disagreement. You may choose to present earlier, historical studies—preferably seminal studies that are of significant importance—and close with most recent work. Another approach is to start with the most distantly related facts and literature and then report on those most closely related to your specific research question. You could also compare and contrast valid approaches, features, characteristics, theories – that is, one approach, then a second approach, followed by a third approach.

Here are some additional tips for writing the body of your literature review:

- Start broad and then narrow down to more specific information.
- When appropriate, cite two or more sources for a single point, but avoid long strings of references for a single point.
- Use quotes sparingly. Quotations for definitions are okay, but reserve quotes for when someone says something so well you couldn't possible phrase it differently. Never use quotes for statistics.
- Paraphrase when you need to relate the specific details within an article, and try to reword it

in a way that is understandable to your audience.
- Include only the aspects of the study that are relevant to your literature review. Don't insert extra facts about a study just to take up space.
- Avoid first-person like language like "I" and "we" to maintain objectivity.
- Avoid informal language like contractions, idioms, and rhetorical questions.
- Note any sections of your review that lack citations and facts from literature. Your arguments need to be based in specific empirical or theoretical facts. Do not approach this like a reflective journal entry.
- Point out consistent findings and emphasize stronger studies over weaker ones.
- Point out important strengths and weaknesses of research studies, as well as contradictions and inconsistent findings.
- Implications and suggestions for further research (where there are gaps in the current literature) should be specific.

The conclusion should summarize your literature review, discuss implications, and create a space for future or further research needed in this area. Your conclusion, like the rest of your literature review, should have a point that you are trying to make. What are the important implications of your literature review? How do they inform the question you are trying to answer?

While you should consult with your professor and their syllabus for the final structure your literature review should take, here is an example of the possible structure for a literature review:

- Problem statement
- Establish the importance of the topic
- Number and type of people affected
- Seriousness of the impact
- Physical, psychological, economic, social consequences of the problem
- Introduction
- Definitions of key terms
- Important arguments you will make
- Overview of the organization of the rest of the review
- Body of the review
- Topic 1
- Supporting evidence
- Topic 2
- Supporting evidence
- Topic 3
- Supporting evidence
- Conclusion
- Implications
- Specific suggestions for future research

- How your research topic adds to the literature

Chapter 4: Conducting a Literature Review is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

Editing Your Literature Review

For your literature review, remember that your goal is to construct an argument for why your research question is interesting and worth addressing—not necessarily why your favorite answer to it is correct. As you start editing your literature review, make sure that it is balanced. If you want to emphasize the generally accepted understanding of a phenomenon, then of course you should discuss various studies that have demonstrated it. However, if there are other studies that have found contradictory findings, you should discuss them, too. Or, if you are proposing a new theory, then you should discuss findings that are consistent with that theory. However, if there are other findings that are inconsistent with it, again, you should discuss them too. It is acceptable to argue that the balance of the research supports the existence of a phenomenon or is consistent with a theory (and that is usually the best that researchers in social work can hope for), but it is not acceptable to ignore contradictory evidence. Besides, a large part of what makes a research question interesting is uncertainty about its answer (University of Minnesota, 2016).

In addition to subjectivity and bias, another obstruction to getting your literature review written is writer's block. Often times, writer's block can come from confusing the creating and editing parts of the writing process. Many writers often start by simply trying to type out what they want to say, regardless of how good it is. First drafts are a natural and important part of the writing process, and they are typically in need of much refinement. Even if you have a detailed outline to work from, the words are not going to fall into place perfectly the first time you start writing. You should consider turning off the editing and critiquing part of your brain for a little while and allow your thoughts to flow. Don't worry about putting the correct internal citation when you first write. Just get the information out. Only after you've reached a natural stopping point might you go back and edit your draft for grammar, APA formatting, organization, flow, and more. Divorcing the writing and editing process can go a long way to addressing writer's block—as can picking a topic about which you have something to say!

As you are editing, keep in mind these questions adapted from Green (2012):

- **Content**: Have I clearly stated the main idea or purpose of the paper and address all the issues? Is the thesis or focus clearly presented and appropriate for the reader?
- **Organization**: How well is it structured? Is the organization spelled out for the reader and easy to follow?
- **Flow**: Is there a logical flow from section to section, paragraph to paragraph, sentence to sentence? Are there transitions between and within paragraphs that link ideas together?
- **Development**: Have I validated the main idea with supporting material? Are supporting data sufficient? Does the conclusion match the introduction?
- **Form**: Are there any APA style issues, redundancy, problematic wording and terminology (always know the definition of any word you use!), flawed sentence constructions and selec-

tion, spelling, and punctuation?

Chapter 4: Conducting a Literature Review is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

References

Bernnard, D., Bobish, G., Hecker, J., Holden, I., Hosier, A., Jacobson, T., Loney, T., & Bullis, D. (2014). Presenting: Sharing what you've learned. In Bobish, G., & Jacobson, T. (eds.) The information literacy users guide: An open online textbook. https://milnepublishing.geneseo.edu/the-information-literacy-users-guide-an-open-online- textbook/chapter/present-sharing-what-youve-learned/

Bhattacherjee, A., (2012). Social science research: Principles, methods, and practices. Textbooks Collection. 3. https://scholarcommons.usf.edu/oa_textbooks/3

Boote, D., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. Educational Researcher 34(6), 3-15.

Burnett, D. (2012). Inscribing knowledge: Writing research in social work. In W. Green & B. L. Simon (Eds.), The Columbia guide to social work writing (pp. 65-82). New York, NY: Columbia University Press.

Green, W. Writing strategies for academic papers. In W. Green & B. L. Simon (Eds.), The Columbia guide to social work writing (pp. 25-47). New York, NY: Columbia University Press.

Houser, J., (2018). Nursing research reading, using, and creating evidence (4th ed.). Burlington, MA: Jones & Bartlett.

Lamott, A. (1995). Bird by bird: Some instructions on writing and life. New York, NY: Penguin.

Loseke, D. (2017). Methodological thinking: Basic principles of social research design (2nd ed.). Los Angeles, CA: Sage.

Machi, L., & McEvoy, B. (2012). The literature review: Six steps to success (2nd ed). Thousand Oaks, CA: Corwin.

O'Gorman, K., & MacIntosh, R. (2015). Research methods for business & management: A guide to writing your dissertation (2nd ed.). Oxford: Goodfellow Publishers.

University of Minnesota Libraries Publishing. (2016). This is a derivative of Research Methods in Psychology by a publisher who has requested that they and the original author not receive attribution, which was originally released and is used under CC BY-NC-SA. This work, unless otherwise expressly stated, is licensed under a Creative Commons Attribution-NonCommercial- ShareAlike 4.0 International License.

Webster, J., & Watson, R. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly, 26(2), xiii-xxiii. https://web.njit.edu/~egan/ Writing_A_Literature_Review.pdf

Chapter 4: Conducting a Literature Review is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

CHAPTER 5: CONCEPTUALIZATION AND OPERATIONALIZATION

In this chapter, we'll discuss measurement, conceptualization, and operationalization. If you're not quite sure what any of those words mean, or even how to pronounce them, no need to worry. By the end of the chapter, you should be able to wow your friends and family with your newfound knowledge of these three difficult to pronounce, but relatively simple to grasp, terms.

LEARNING OBJECTIVES

- Define measurement.
- Describe Kaplan's three categories of the things that social scientists measure.
- Identify the stages at which measurement is important.
- Define concept.
- Describe why defining our concepts is important.
- Describe how conceptualization works.
- Define dimensions in terms of social scientific measurement.
- Describe reification.
- Describe how operationalization works.
- Define and give an example of indicators.
- Define reliability and give examples of different types
- Define validity and give examples of different types
- Define and provide examples for each of the four levels of measurement.

Measurement

Measurement is important. Recognizing that fact, and respecting it, will be of great benefit to you—both in research methods and in other areas of life as well. If, for example, you have ever baked a cake, you know well the importance of measurement. As someone who much prefers rebelling against precise rules over following them, I once learned the hard way that measurement matters. A couple of years ago I attempted to bake my wife a birthday cake without the help of any measuring utensils. I'd baked before, I reasoned, and I had a pretty good sense of the difference between a cup and a tablespoon. How hard could it be? As it turns out, it's not easy guesstimating precise measures. The cake was inedible.

Just as measurement is critical to successful baking, it is as important to successfully pulling off a social scientific research project. In sociology, when we use the term measurement we mean the process by which we describe and ascribe meaning to the key facts, concepts, or other phenomena that we are investigating. At its core, measurement is about defining one's terms in as clear and precise a way as possible. Of course, measurement in social science isn't quite as simple as using some predetermined or universally agreed-on tool, such as a measuring cup or spoon, but there are some basic tenants on which most social scientists agree when it comes to measurement. We'll explore those as well as some of the ways that measurement might vary depending on your unique approach to the study of your topic.

What Do Social Scientists Measure?

The question of what social scientists measure can be answered by asking oneself what social scientists study. Think about the topics you've learned about in other sociology classes you've taken or the topics you've considered investigating yourself. Or think about the many examples of research you've read about in this text. In we learned about Melissa Milkie and Catharine Warner's study (2011) [1] of first graders' mental health. In order to conduct that study, Milkie and Warner needed to have some idea about how they were going to measure mental health. What does mental health mean, exactly? And how do we know when we're observing someone whose mental health is good and when we see someone whose mental health is compromised? Understanding how measurement works in research methods helps us answer these sorts of questions.

As you might have guessed, social scientists will measure just about anything that they have an interest in investigating. For example, those who are interested in learning something about the correlation between social class and levels of happiness must develop some way to measure both social class and happiness. Those who wish to understand how well immigrants cope in their new locations must measure immigrant status and coping.

Those who wish to understand how a person's gender shapes their workplace experiences must measure gender and workplace experiences. You get the idea. Social scientists can and do measure just about anything you can imagine observing or wanting to study. Of course, some things are easier to observe, or measure, than others, and the things we might wish to measure don't necessarily all fall into the same category of measureables.

In 1964, philosopher Abraham Kaplan (1964) [2] wrote what has since become a classic work in research methodology, The Conduct of Inquiry (Babbie, 2010). [3] In his text, Kaplan describes different categories of things that behavioral scientists observe. One of those categories, which Kaplan called "observational terms," is probably the simplest to measure in social science. Observational terms are the sorts of things that we can see with the naked eye simply by looking at them. They are terms that "lend themselves to easy and confident verification" (1964, p. 54). [4] If, for example, we wanted to know how the conditions of playgrounds differ across different neighborhoods, we could directly observe the variety, amount, and condition of equipment at various playgrounds.

Indirect observables, on the other hand, are less straightforward to assess. They are "terms whose application calls for relatively more subtle, complex, or indirect observations, in which inferences play an acknowledged part. Such inferences concern presumed connections, usually causal, between what is directly observed and what the term signifies" (1964, p. 55). [5] If we conducted a study for which we wished to know a person's income, we'd probably have to ask them their income, perhaps in an interview or a survey. Thus we have observed income, even if it has only been observed indirectly. Birthplace might be another indirect observable. We can ask study participants where they were born, but chances are good we won't have directly observed any of those people being born in the locations they report.

Sometimes the measures that we are interested in are more complex and more abstract than observational terms or indirect observables. Think about some of the concepts you've learned about in other sociology classes— ethnocentrism, for example. What is ethnocentrism? Well, you might know from your intro to sociology class that it has something to do with the way a person judges another's culture. But how would you measure it?

Here's another construct: bureaucracy. We know this term has something to do with organizations and how they operate, but measuring such a construct is trickier than measuring, say, a person's income. In both cases, ethnocentrism and bureaucracy, these theoretical notions represent ideas whose meaning we have come to agree on. Though we may not be able to observe these abstractions directly, we can observe the confluence of things that they are made up of. Kaplan

referred to these more abstract things that behavioral scientists measure as constructs. Constructs are "not observational either directly or indirectly" (1964, p. 55), [6] but they can be defined based on observables.

Thus far we have learned that social scientists measure what Abraham Kaplan called observational terms, indirect observables, and constructs. These terms refer to the different sorts of things that social scientists may be interested in measuring. But how do social scientists measure these things? That is the next question we'll tackle.

How Do Social Scientists Measure?

Measurement in social science is a process. It occurs at multiple stages of a research project: in the planning stages, in the data collection stage, and sometimes even in the analysis stage. Recall that previously we defined measurement as the process by which we describe and ascribe meaning to the key facts, concepts, or other phenomena that we are investigating. Once we've identified a research question, we begin to think about what some of the key ideas are that we hope to learn from our project. In describing those key ideas, we begin the measurement process.

Let's say that our research question is the following: How do new university students cope with the adjustment to university? In order to answer this question, we'll need to some idea about what coping means. We may come up with an idea about what coping means early in the research process, as we begin to think about what to look for (or observe) in our data-collection phase. Once we've collected data on coping, we also have to decide how to report on the topic. Perhaps, for example, there are different types or dimensions of coping, some of which lead to more successful adjustment than others. However we decide to proceed, and whatever we decide to report, the point is that measurement is important at each of these phases.

As the preceding paragraph demonstrates, measurement is a process in part because it occurs at multiple stages of conducting research. We could also think of measurement as a process because of the fact that measurement in itself involves multiple stages. From identifying one's key terms to defining them to figuring out how to observe them and how to know if our observations are any good, there are multiple steps involved in the measurement process. An additional step in the measurement process involves deciding what elements one's measures contain. A measure's elements might be very straightforward and clear, particularly if they are directly observable. Other measures are more complex and might require the researcher to account for different themes or types. These sorts of complexities require paying careful attention to a concept's level of measurement and its dimensions. We'll explore these complexities in greater depth at the end of this chapter, but first let's look more closely at the early steps involved in the measurement process.

Conceptualization

In this section we'll take a look at one of the first steps in the measurement process, conceptualization. This has to do with defining our terms as clearly as possible and also not taking ourselves too seriously in the process. Our definitions mean only what we say they mean—nothing more and nothing less. Let's talk first about how to define our terms, and then we'll examine what I mean about not taking ourselves (or our terms, rather) too seriously.

Concepts and Conceptualization

So far, the word concept has come up quite a bit, and it would behoove us to make sure we have a shared understanding of that term. A concept is the notion or image that we conjure up when we think of some cluster of related observations or ideas. For example, masculinity is a concept. What do you think of when you hear that word? Presumably, you imagine some set of behaviors and perhaps even a particular style of self-presentation. Of course, we can't necessarily assume that everyone conjures up the same set of ideas or images when they hear the word masculinity. In fact, there are many possible ways to define the term. While some definitions may be more common or have more support than others, there isn't one true, always-correct-in-all-settings definition. What counts as masculine may shift over time, from culture to culture, and even from individual to individual (Kimmel, 2008). [1] This is why defining our concepts is so important.

You might be asking yourself why you should bother defining a term for which there is no single, correct definition. Believe it or not, this is true for any concept you might measure in a social scientific study—there is never a single, always-correct definition. When we conduct empirical research, our terms mean only what we say they mean—nothing more and nothing less. There's a New Yorker cartoon that aptly represents this idea (https://www.cartoonbank.com/1998/it-all-depends-on-how-you-define-chop/invt/117721). It depicts a young George Washington holding an ax and standing near a freshly chopped cherry tree. Young George is looking up at a frowning adult who is standing over him, arms crossed. The caption depicts George explaining, "It all depends on how you define 'chop.'" Young George Washington gets the idea—whether he actually chopped down the cherry tree depends on whether we have a shared understanding of the term chop.

Without a shared understanding of this term, our understandings of what George has just done may differ. Likewise, without understanding how a researcher has defined her or his key concepts, it would be nearly impossible to understand the meaning of that researcher's findings and conclu-

sions. Thus any decision we make based on findings from empirical research should be made based on full knowledge not only of how the research was designed but also of how its concepts were defined and measured.

So how do we define our concepts? This is part of the process of measurement, and this portion of the process is called conceptualization. ***Conceptualization*** involves writing out clear, concise definitions for our key concepts. Sticking with the previously mentioned example of masculinity, think about what comes to mind when you read that term. How do you know masculinity when you see it? Does it have something to do with men? With social norms? If so, perhaps we could define masculinity as the social norms that men are expected to follow. That seems like a reasonable start, and at this early stage of conceptualization, brainstorming about the images conjured up by concepts and playing around with possible definitions is appropriate. But this is just the first step. It would make sense as well to consult other previous research and theory to understand if other scholars have already defined the concepts we're interested in. This doesn't necessarily mean we must use their definitions, but understanding how concepts have been defined in the past will give us an idea about how our conceptualizations compare with the predominant ones out there. Understanding prior definitions of our key concepts will also help us decide whether we plan to challenge those conceptualizations or rely on them for our own work.

If we turn to the literature on masculinity, we will surely come across work by Michael Kimmel, one of the preeminent masculinity scholars in the United States. After consulting Kimmel's prior work (2000; 2008), we might tweak our initial definition of masculinity just a bit. Rather than defining masculinity as "the social norms that men are expected to follow," perhaps instead we'll define it as "the social roles, behaviors, and meanings prescribed for men in any given society at any one time." Our revised definition is both more precise and more complex. Rather than simply addressing one aspect of men's lives (norms), our new definition addresses three aspects: roles, behaviors, and meanings. It also implies that roles, behaviors, and meanings may vary across societies and over time. Thus, to be clear, we'll also have to specify the particular society and time period we're investigating as we conceptualize masculinity.

As you can see, conceptualization isn't quite as simple as merely applying any random definition that we come up with to a term. Sure, it may involve some initial brainstorming, but conceptualization goes beyond that. Once we've brainstormed a bit about the images a particular word conjures up for us, we should also consult prior work to understand how others define the term in question. And after we've identified a clear definition that we're happy with, we should make sure that every term used in our definition will make sense to others. Are there terms used within our definition that also need to be defined? If so, our conceptualization is not yet complete. And there is yet another aspect of conceptualization to consider: concept dimensions. We'll consider that aspect along with an additional word of caution about conceptualization next.

A Word of Caution About Conceptualization

So now that we've come up with a clear definition for the term masculinity and made sure that the terms we use in our definition are equally clear, we're done, right? Not so fast. If you've ever met more than one man in your life, you've probably noticed that they are not all exactly the same, even if they live in the same society and at the same historical time period. This could mean that there are dimensions of masculinity. In terms of social scientific measurement, concepts can be said to have ***dimensions*** when there are multiple elements that make up a single concept. With respect to the term masculinity, dimensions could be regional (Is masculinity defined differently in different regions of the same country?), age based (Is masculinity defined differently for men of different ages?), or perhaps power based (Are some forms of masculinity valued more than others?). In any of these cases, the concept masculinity would be considered to have multiple dimensions. While it isn't necessarily a must to spell out every possible dimension of the concepts you wish to measure, it may be important to do so depending on the goals of your research. The point here is to be aware that some concepts have dimensions and to think about whether and when dimensions may be relevant to the concepts you intend to investigate.

Before we move on to the additional steps involved in the measurement process, it would be wise to caution ourselves about one of the dangers associated with conceptualization. While I've suggested that we should consult prior scholarly definitions of our concepts, it would be wrong to assume that just because prior definitions exist that they are any more real than whatever definitions we make up (or, likewise, that our own made-up definitions are any more real than any other definition). It would also be wrong to assume that just because definitions exist for some concept that the concept itself exists beyond some abstract idea in our heads. This idea, assuming that our abstract concepts exist in some concrete, tangible way, is known as reification.

To better understand reification, take a moment to think about the concept of social structure. This concept is central to sociological thinking. When we social scientists talk about social structure, we are talking about an abstract concept. Social structures shape our ways of being in the world and of interacting with one another, but they do not exist in any concrete or tangible way. A social structure isn't the same thing as other sorts of structures, such as buildings or bridges. Sure, both types of structures are important to how we live our everyday lives, but one we can touch, and the other is just an idea that shapes our way of living.

Here's another way of thinking about reification: Think about the term family. If you were interested in studying this concept, we've learned that it would be good to consult prior theory and research to understand how the term has been conceptualized by others. But we should also question past conceptualizations. Think, for example, about where we'd be today if we used the same definition of family that was used, say, 100 years ago. How have our understandings of this concept changed over time? What role does conceptualization in social scientific research play in our cultural understandings of terms like family? The point is that our terms mean nothing more and

nothing less than whatever definition we assign to them. Sure, it makes sense to come to some social agreement about what various concepts mean. Without that agreement, it would be difficult to navigate through everyday living. But at the same time, we should not forget that we have assigned those definitions and that they are no more real than any other, alternative definition we might choose to assign.

Operationalization

Now that we have figured out how to define, or conceptualize, our terms we'll need to think about operationalizing them. ***Operationalization*** is the process by which we spell out precisely how a concept will be measured. It involves identifying the specific research procedures we will use to gather data about our concepts. This of course requires that one know what research method(s) he or she will employ to learn about her or his concepts, and we'll examine specific research methods in through . For now, let's take a broad look at how operationalization works. We can then revisit how this process works when we examine specific methods of data collection in later chapters.

Indicators

Operationalization works by identifying specific ***indicators*** that will be taken to represent the ideas that we are interested in studying. If, for example, we are interested in studying masculinity, indicators for that concept might include some of the social roles prescribed to men in society such as breadwinning or fatherhood. Being a breadwinner or a father might therefore be considered indicators of a person's masculinity. The extent to which a man fulfills either, or both, of these roles might be understood as clues (or indicators) about the extent to which he is viewed as masculine.

Let's look at another example of indicators. Each day, Gallup researchers poll 1,000 randomly selected Americans to ask them about their well-being. To measure well-being, Gallup asks these people to respond to questions covering six broad areas: physical health, emotional health, work environment, life evaluation, healthy behaviors, and access to basic necessities. Gallup uses these six factors as indicators of the concept that they are really interested in: well-being (https://www.gallup.com/poll/123215/Gallup-Healthways- Index.aspx).

Identifying indicators can be even simpler than the examples described thus far. What are the possible indicators of the concept of gender? Most of us would probably agree that "woman" and "man" are both reasonable indicators of gender, and if you're a social scientist of gender, you might also add an indicator of "other" to the list. Political party is another relatively easy concept for which to identify indicators. In the United States, likely indicators include Democrat and Republican and, depending on your research interest, you may include additional indicators such as Independent, Green, or Libertarian as well. Age and birthplace are additional examples of concepts for which identifying indicators is a relatively simple process. What concepts are of interest to you, and what are the possible indictors of those concepts?

We have now considered a few examples of concepts and their indicators but it is important that we don't make the process of coming up with indicators too arbitrary or casual. One way to avoid taking an overly casual approach in identifying indicators, as described previously, is to turn to prior theoretical and empirical work in your area. Theories will point you in the direction of relevant concepts and possible indicators; empirical work will give you some very specific examples of how the important concepts in an area have been measured in the past and what sorts of indicators have been used. Perhaps it makes sense to use the same indicators as researchers who have come before you. On the other hand, perhaps you notice some possible weaknesses in measures that have been used in the past that your own methodological approach will enable you to overcome. Speaking of your methodological approach, another very important thing to think about when deciding on indicators and how you will measure your key concepts is the strategy you will use for data collection. A survey implies one way of measuring concepts, while field research implies a quite different way of measuring concepts. Your data-collection strategy will play a major role in shaping how you operationalize your concepts.

Putting It All Together

Moving from identifying concepts to conceptualizing them and then to operationalizing them is a matter of increasing specificity. You begin with a general interest, identify a few concepts that are essential for studying that interest, work to define those concepts, and then spell out precisely how you will measure those concepts. Your focus becomes narrower as you move from a general interest to operationalization. One point not yet mentioned is that while the measurement process often works as outlined above, it doesn't necessarily always have to work out that way. What if your interest is in discovering how people define the same concept differently? If that's the case, you probably begin the measurement process the same way as outlined earlier, by having some general interest and identifying key concepts related to that interest. You might even have some working definitions of the concepts you wish to measure. And of course you'll have some idea of how you'll go about discovering how your concept is defined by different people. But you may not go so far as to have a clear set of indicators identified before beginning data collection, for that would defeat the purpose if your aim is to discover the variety of indicators people rely on.

Let's consider an example of when the measurement process may not work out exactly as depicted above. Blackstone (2003) conducted a study to compare activism in the breast cancer movement with activism in the anti-rape movement. A goal of this study was to understand what "politics" means in the context of social movement participation. By observing participants to understand how they engaged in politics, an understanding was developed of what politics meant for these groups and individuals: politics seemed to be about power, "who has it, who wants it, and how it is given, negotiated and taken away" (Blackstone, 2007). Specific actions, such as the awareness-raising bicycle event Ride Against Rape, seemed to be political in that they empowered survivors to see that they were not alone, and they empowered clinics (through funds raised at the event) to provide services to survivors. By taking the time to observe movement participants in action for many months, Blackstone was able to learn how politics operated in the day-to-day goings-on of social movements and in the lives of movement participants. While it was not evident at the outset of the study, observations led to defining politics as linked to action and challenging power. In this case, observations *preceded* coming up with a clear definition for my key term, and certainly before identifying indicators for the term. The measurement process therefore worked more inductively than implied that it might.

Measurement Quality

In quantitative research, once we've managed to define our terms and specify the operations for measuring them, how do we know that our measures are any good? Without some assurance of the quality of our measures, we cannot be certain that our findings have any meaning or, at the least, that our findings mean what we think they mean. When social scientists measure concepts, they aim to achieve reliability and validity in their measures. These two aspects of measurement quality are the focus of this section. We'll consider reliability first and then take a look at validity. For both aspects of measurement quality, let's say our interest is in measuring the concepts of alcoholism and alcohol intake. What are some potential problems that could arise when attempting to measure this concept, and how might we work to overcome those problems?

Reliability

First, let's say we've decided to measure alcoholism by asking people to respond to the following question: Have you ever had a problem with alcohol? If we measure alcoholism in this way, it seems likely that anyone who identifies as an alcoholic would respond with a yes to the question. So, this must be a good way to identify our group of interest, right? Well, maybe. Think about how you or others you know would respond to this question. Would responses differ after a wild night out from what they would have been the day before? Might an infrequent drinker's current headache from the single glass of wine she had last night influence how she answers the question this morning? How would that same person respond to the question before consuming the wine? In each of these cases, if the same person would respond differently to the same question at different points, it is possible that our measure of alcoholism has a reliability problem. Reliability in measurement is about consistency.

One common problem of reliability with social scientific measures is memory. If we ask research participants to recall some aspect of their own past behavior, we should try to make the recollection process as simple and straightforward for them as possible. Sticking with the topic of alcohol intake, if we ask respondents how much wine, beer, and liquor they've consumed each day over the course of the past 3 months, how likely are we to get accurate responses? Unless a person keeps a journal documenting their intake, there will very likely be some inaccuracies in their responses. If, on the other hand, we ask a person how many drinks of any kind they have consumed in the past week, we might get a more accurate set of responses.

Reliability can be an issue even when we're not reliant on others to accurately report their behaviors. Perhaps a researcher is interested in observing how alcohol intake influences interactions in public locations. She may decide to conduct observations at a local pub, noting how many drinks patrons consume and how their behavior changes as their intake changes. But what if the researcher has to use the restroom and misses the three shots of tequila that the person next to her downs during the brief period she is away? The reliability of this researcher's measure of alcohol intake, counting numbers of drinks she observes patrons consume, depends on her ability to actually observe every instance of patrons consuming drinks. If she is unlikely to be able to observe every such instance, then perhaps her mechanism for measuring this concept is not reliable.

If a measure is ***reliable***, it means that if the measure is given multiple times, the results will be consistent each time. For example, if you took the SATs on multiple occasions before coming to school, your scores should be relatively the same from test to test. This is what is known as ***test-retest reliability***. In the same way, if a person is clinically depressed, a depression scale should give similar (though not necessarily identical) results today that it does two days from now.

Additionally, if your study involves observing people's behaviors, for example watching sessions of mothers playing with infants, you may also need to assess ***inter-rater reliability***. Inter-rater reliability is the degree to which different observers agree on what happened. Did you miss when the infant offered an object to the mother and the mother dismissed it? Did the other person rating miss that event? Do you both similarly rate the parent's engagement with the child? Again, scores of multiple observers should be consistent, though perhaps not perfectly identical.

Finally, for scales, ***internal consistency reliability*** is an important concept. The scores on each question of a scale should be correlated with each other, as they all measure parts of the same concept. Think about a scale of depression, like Beck's Depression Inventory. A person who is depressed would score highly on most of the measures, but there would be some variation. If we gave a group of people that scale, we would imagine there should be a correlation between scores on, for example, mood disturbance and lack of enjoyment. They aren't the same concept, but they are related. So, there should be a mathematical relationship between them. A specific statistical test known as Cronbach's Alpha provides a way to measure how well each question of a scale is related to the others.

Test-retest, inter-rater, and internal consistency are three important subtypes of reliability. Researchers use these types of reliability to make sure their measures are consistently measuring the concepts in their research questions.

Validity

While reliability is about consistency, ***validity*** is about accuracy. What image comes to mind for you when you hear the word alcoholic? Are you certain that the image you conjure up is similar to the image others have in mind? If not, then we may be facing a problem of validity.

For a measure to have validity, we must be certain that our measures accurately get at the meaning of our concepts. Think back to the first possible measure of alcoholism we considered in the previous few paragraphs. There, we initially considered measuring alcoholism by asking research participants the following question: Have you ever had a problem with alcohol? We realized that this might not be the most reliable way of measuring alcoholism because the same person's response might vary dramatically depending on how they are feeling that day. Likewise, this measure of alcoholism is not particularly valid. What is "a problem" with alcohol? For some, it might be having had a single regrettable or embarrassing moment that resulted from consuming too much. For others, the threshold for "problem" might be different; perhaps a person has had numerous embarrassing drunken moments but still gets out of bed for work every day, so they don't perceive themselves as having a problem. Because what each respondent considers to be problematic could vary so dramatically, our measure of alcoholism isn't likely to yield any useful or meaningful results if our aim is to objectively understand, say, how many of our research participants are alcoholics.

In the last paragraph, critical engagement with our measure for alcoholism "Do you have a problem with alcohol?" was shown to be flawed. We assessed its ***face validity*** or whether it is plausible that the question measures what it intends to measure. Face validity is a subjective process. Sometimes face validity is easy, as a question about height wouldn't have anything to do with alcoholism. Other times, face validity can be more difficult to assess. Let's consider another example.

Perhaps we're interested in learning about a person's dedication to healthy living. Most of us would probably agree that engaging in regular exercise is a sign of healthy living, so we could measure healthy living by counting the number of times per week that a person visits their local gym. But perhaps they visit the gym to use their tanning beds or to flirt with potential dates or sit in the sauna. These activities, while potentially relaxing, are probably not the best indicators of healthy living. Therefore, recording the number of times a person visits the gym may not be the most valid way to measure their dedication to healthy living.

Another problem with this measure of healthy living is that it is incomplete. ***Content validity*** assesses for whether the measure includes all of the possible meanings of the concept. Think back to the previous section on multidimensional variables. Healthy living seems like a multidimensional concept that might need an index, scale, or typology to measure it completely. Our one question on gym attendance doesn't cover all aspects of healthy living. Once you have created one, or found one in the existing literature, you need to assess for content validity. Are there other aspects of healthy living that aren't included in your measure?

Let's say you have created (or found) a good scale for your measure of healthy living. A valid measure of healthy living would be able to predict, for example, scores of a blood panel test during their annual physical. This is called ***predictive validity***, and it means that your measure predicts things it should be able to predict. In this case, I assume that if you have a healthy lifestyle, a standard blood test done a few months later during an annual checkup would show healthy results. If we were to administer the blood panel measure at the same time as you administer your scale of healthy living, we would be assessing concurrent validity. Concurrent validity is the same as predictive validity—the scores on your measure should be similar to an established measure—except that both measures are given at the same time.

Another closely related concept is ***convergent validity***. In assessing for convergent validity, one should look for existing measures of the same concept, for example the Healthy Lifestyle Behaviors Scale (HLBS). If you give someone your scale and the HLBS at the same time, their scores should be pretty similar. Convergent validity takes an existing measure of the same concept and compares your measure to it. If their scores are similar, then it's probably likely that they are both measuring the same concept. Discriminant validity is a similar concept, except you would be comparing your measure to one that is entirely unrelated. A participant's scores on your healthy lifestyle measure shouldn't be statistically correlated with a scale that measures knowledge of the Italian language.

These are the basic subtypes of validity, though there are certainly others you can read more about. One way to think of validity is to think of it as you would a portrait. Some portraits of people look just like the actual person they are intended to represent. But other representations of people's images, such as caricatures and stick drawings, are not nearly as accurate. While a portrait may not be an exact representation of how a person looks, what's important is the extent to which it approximates the look of the person it is intended to represent. The same goes for validity in measures. No measure is exact, but some measures are more accurate than others.

Complexities in Measurement

You should now have some idea about how conceptualization and operationalization work, and you also know a bit about how to assess the quality of your measures. But measurement is sometimes a complex process, and some concepts are more complex than others. Measuring a person's political party affiliation, for example, is less complex than measuring her or his sense of alienation. In this section we'll consider some of these complexities in measurement. First, we'll take a look at the various levels of measurement that exist, and then we'll consider a couple strategies for capturing the complexities of the concepts we wish to measure.

Levels of Measurement

When social scientists measure concepts, they sometimes use the language of variables and attributes. A variable refers to a grouping of several characteristics. Attributes are those characteristics. A variable's attributes determine its level of measurement. There are four possible levels of measurement; they are ***nominal, ordinal, interval, and ratio***.

At the ***nominal*** level of measurement, variable attributes meet the criteria of exhaustiveness and mutual exclusivity. This is the most basic level of measurement. Relationship status, gender, race, political party affiliation, and religious affiliation are all examples of nominal-level variables. For example, to measure relationship status, we might ask respondents to tell us if they are currently partnered or single. These two attributes pretty much exhaust the possibilities for relationship status (i.e., everyone is always one or the other of these), and it is not possible for a person to simultaneous occupy more than one of these statuses (e.g., if you are single, you cannot also be partnered). Thus this measure of relationship status meets the criteria that nominal- level attributes must be ***exhaustive*** and ***mutually exclusive***. One unique feature of nominal-level measures is that they cannot be mathematically quantified. We cannot say, for example, that being partnered has more or less quantifiable value than being single (note we're not talking here about the economic impact of one's relationship status—we're talking only about relationship status on its own, not in relation to other variables).

Unlike nominal-level measures, attributes at the ***ordinal*** level can be rank ordered, though we cannot calculate a mathematical distance between those attributes. We can simply say that one attribute of an ordinal-level variable is more or less than another attribute. Ordinal-level attributes are also exhaustive and mutually exclusive, as with nominal-level variables. Examples of ordinal-level measures include social class, degree of support for policy initiatives, television program

rankings, and prejudice. Thus while we can say that one person's support for some public policy may be more or less than his neighbor's level of support, we cannot say exactly how much more or less.

At the ***interval*** level, measures meet all the criteria of the two preceding levels, plus the distance between attributes is known to be equal. IQ scores are interval level, as are temperatures. Interval-level variables are not particularly common in social science research, but their defining characteristic is that we can say how much more or less one attribute differs from another. We cannot, however, say with certainty what the ratio of one attribute is in comparison to another. For example, it would not make sense to say that 50 degrees is half as hot as 100 degrees.

Finally, at the ***ratio*** level, attributes are mutually exclusive and exhaustive, attributes can be rank ordered, the distance between attributes is equal, and attributes have a true zero point. Thus with these variables, we can say what the ratio of one attribute is in comparison to another. Examples of ratio-level variables include age and years of education. We know, for example, that a person who is 12 years old is twice as old as someone who is 6 years old.

Key Takeaways, Exercises, and References

Key Takeaways

- Measurement is the process by which we describe and ascribe meaning to the key facts, concepts, or other phenomena that we are investigating.
- Kaplan identified three categories of things that social scientists measure including observational terms, indirect observables, and constructs.
- Measurement occurs at all stages of research.
- Conceptualization is a process that involves coming up with clear, concise definitions.
- Some concepts have multiple elements or dimensions.
- Just because definitions for abstract concepts exist does not mean that the concept is tangible or concrete.
- Operationalization involves spelling out precisely how a concept will be measured.
- The measurement process generally involves going from a more general focus to a narrower one, but the process does not proceed in exactly the same way for all research projects.
- Reliability is a matter of consistency.
- Validity is a matter of accuracy.
- There are many types of validity and reliability.
- In social science, our variables can be one of four different levels of measurement: nominal, ordinal, interval, or ratio.

Exercises

- See if you can come up with one example of each of the following: an observational term, an indirect observable, and a construct. How might you measure each?
- Conceptualize the term discipline and identify possible dimensions of the term. Have someone who is in the class with you do the same thing (without seeing your conceptualization). Now compare what you each came up with. How do your conceptualizations and dimensions differ, and why?
- Identify a concept that is important in your area of interest. Challenge yourself to conceptualize the term without first consulting prior literature. Now consult prior work to see how your concept has been conceptualized by others. How and where does your conceptualization dif-

fer from others? Are there dimensions of the concept that you or others hadn't considered?
- Think of a concept that is of interest to you. Now identify some possible indicators of that concept.
- Operationalize a concept that is of interest to you. What are some possible problems of reliability or validity that you could run into given your operationalization? How could you tweak your operationalization and overcome those problems?
- Sticking with the same concept you identified in exercise 1, find out how other social scientists have operationalized this concept. You can do this by revisiting readings from other sociology courses you've taken or by looking up a few articles using Sociological Abstracts. How does your plan for operationalization differ from that used in previous research? What potential problems of reliability or validity do you see? How do the researchers address those problems?
- Together with a fellow research methods student, identify six concepts that are of interest to you both. , on your own, identify each concept's level of measurement. Share your answers with your peer. Discuss why you chose each level of measurement that you chose and, together, try to come to some agreement about any concepts that you labeled differently.
- Take a look at Gallup's page on their well-being index: https://www.gallup.com/poll/123215/Gallup-Healthways-Index.aspx. Read about how various concepts there are operationalized and indexed.

References

Babbie, E. (2010). The practice of social research (12th ed.). Belmont, CA: Wadsworth.

Blackstone, A. (2003). Racing for the cure and taking back the night: Constructing gender, politics, and public participation in women's activist/volunteer work. PhD dissertation, Department of Sociology, University of Minnesota, Minneapolis, MN.

Blackstone, A. (2007). Finding politics in the silly and the sacred: Anti-rape activism on campus. Sociological Spectrum, 27, 151–163.

Blackstone, A., & McLaughlin, H. (2009). Sexual harassment. In J. O'Brien & E. L. Shapiro (Eds.), Encyclopedia of gender and society (pp. 762– 766). Thousand Oaks, CA: Sage.

Durkheim, E. (1897 [2006 translation by R. Buss]). On suicide. London, UK: Penguin.

Kaplan, A. (1964). The conduct of inquiry: Methodology for behavioral science. San Francisco, CA: Chandler Publishing Company, p. 54.

Kaplan, A. (1964). The conduct of inquiry: Methodology for behavioral science. San Francisco, CA: Chandler Publishing Company, p. 55.

Kaplan, A. (1964). The conduct of inquiry: Methodology for behavioral science. San Francisco, CA: Chandler Publishing Company, p. 55.

Kaplan, A. (1964). The conduct of inquiry: Methodology for behavioral science. San Francisco, CA: Chandler Publishing Company.

Kimmel, M. (2000). The gendered society. New York, NY: Oxford University Press; Kimmel, M. (2008). Masculinity. In W. A. Darity Jr. (Ed.), International encyclopedia of the social sciences (2nd ed., Vol. 5, pp. 1–5). Detroit, MI: Macmillan Reference USA.

Kimmel, M. (2008). Masculinity. In W. A. Darity Jr. (Ed.), International encyclopedia of the social sciences (2nd ed., Vol. 5, pp. 1–5). Detroit, MI: Macmillan Reference USA.

Milkie, M. A., & Warner, C. H. (2011). Classroom learning environments and the mental health of first grade children. Journal of Health and Social Behavior, 52, 4–22.

CHAPTER 6: SAMPLING

If you were to see a cute photo of babies hanging out together and one of them was wearing a green onesie, if you were to conclude that all babies wore green based on the photo that you would have committed selective observation. In that example of informal observation, our sampling strategy (just observing the baby in green) was of course faulty, but we nevertheless have engaged in sampling. Sampling has to do with selecting some subset of one's group of interest (in this case, babies) and drawing conclusions from that subset. How we sample and who we sample shapes what sorts of conclusions we are able to draw. Ultimately, this chapter focuses on questions about who or what you want to be able to make claims about in your research. In the following sections, we'll define sampling, discuss different types of sampling strategies, and consider how to judge the quality of samples as consumers of social scientific research.

LEARNING OBJECTIVES

- Understand the difference between populations and samples.
- Define nonprobability sampling, and describe instances in which a researcher might choose a nonprobability sampling technique.
- Describe the different types of nonprobability samples.
- Describe how probability sampling differs from nonprobability sampling.
- Define generalizability, and describe how it is achieved in probability samples.
- Identify the various types of probability samples, and provide a brief description of each.
- Identify several questions we should ask about samples when reading the results of research.
- Name some tenets worth keeping in mind with respect to responsibly reading research findings.

Population Versus Samples

When I teach research methods, my students are sometimes disheartened to discover that the research projects they complete during the course will not make it possible for them to make sweeping claims about "all" of whomever it is that they're interested in studying. What they fail to realize, however, is that they are not alone. One of the most surprising and frustrating lessons research methods students learn is that there is a difference between one's population of interest and one's study sample. While there are certainly exceptions, more often than not a researcher's population and her or his sample are not the same.

In social scientific research, a ***population*** is the cluster of people, events, things, or other phenomena that you are most interested in; it is often the "who" or "what" that you want to be able to say something about at the end of your study. Populations in research may be rather large, such as "the American people," but they are more typically a little less vague than that. For example, a large study for which the population of interest really is the American people will likely specify which American people, such as adults over the age of 18 or citizens or legal residents. A ***sample***, on the other hand, is the cluster of people or events, for example, from or about which you will actually gather data. Some sampling strategies allow researchers to make claims about populations that are much larger than their actual sample with a fair amount of confidence. Other sampling strategies are designed to allow researchers to make theoretical contributions rather than to make sweeping claims about large populations. We'll discuss both types of strategies later in this chapter.

It is quite rare for a researcher to gather data from their entire population of interest. This might sound surprising or disappointing until you think about the kinds of research questions that social scientists typically ask. For example, let's say we wish to answer the following research question: "How do men's and women's university experiences differ, and how are they similar?" Would you expect to be able to collect data from all university students across all nations from all historical time periods? Unless you plan to make answering this research question your entire life's work (and then some), I'm guessing your answer is a resounding no way. So what to do? Does not having the time or resources to gather data from every single person of interest mean having to give up your research interest? Absolutely not. It just means having to make some hard choices about sampling, and then being honest with yourself and your readers about the limitations of your study based on the sample from whom you were able to actually collect data.

Sampling is the process of selecting observations that will be analyzed for research purposes. Both qualitative and quantitative researchers use sampling techniques to help them identify the what or whom from which they will collect their observations. Because the goals of qualitative and quantitative research differ, however, so, too, do the sampling procedures of the researchers employing these methods. First, we examine sampling types and techniques used in qualitative research. After that, we'll look at how sampling typically works in quantitative research.

Sampling Without Generalizing

Qualitative researchers are not as concerned about generalizing to broader populations, but typically make sampling choices that enable them to deepen understanding of whatever phenomenon it is that they are studying. In this section we'll examine the strategies that do not allow for generalizations beyond the sample—used by both qualitative and quantitative researchers in certain instances.

Nonprobability Sampling

Nonprobability sampling refers to sampling techniques for which a person's (or event's or researcher's focus's) likelihood of being selected for membership in the sample is unknown. Because we don't know the likelihood of selection, we don't know with nonprobability samples whether a sample represents a larger population or not. That's OK, though, because representing the population is not the goal with nonprobability samples. That said, the fact that nonprobability samples do not represent a larger population does not mean that they are drawn arbitrarily or without any specific purpose in mind. In the following subsection, "Types of Nonprobability Samples," we'll take a closer look at the process of selecting research elements when drawing a nonprobability sample. But first, let's consider why a researcher might choose to use a nonprobability sample.

So when are nonprobability samples ideal? One instance might be when we're designing a research project. For example, if we're conducting survey research, we may want to administer our survey to a few people who seem to resemble the folks we're interested in studying in order to help work out kinks in the survey. We might also use a nonprobability sample at the early stages of a research project, if we're conducting a pilot study or some exploratory research. This can be a quick way to gather some initial data and help us get some idea of the lay of the land before conducting a more extensive study. From these examples, we can see that nonprobability samples can be useful for setting up, framing, or beginning research. But it isn't just early stage research that relies on and benefits from nonprobability sampling techniques.

Researchers also use nonprobability samples in full-blown research projects. These projects are usually qualitative in nature, where the researcher's goal is in-depth, idiographic understanding rather than more general, nomothetic understanding. Evaluation researchers whose aim is to describe some very specific small group might use nonprobability sampling techniques, for example. Researchers interested in contributing to our theoretical understanding of some phenomenon might also collect data from nonprobability samples. Thus researchers interested in contributing

to social theories, by either expanding on them, modifying them, or poking holes in their propositions, may use nonprobability sampling techniques to seek out cases that seem anomalous in order to understand how theories can be improved.

In sum, there are a number and variety of instances in which the use of nonprobability samples makes sense. We'll examine several specific types of nonprobability samples in the next subsection.

Types of Nonprobability Samples

There are several types of nonprobability samples that researchers use. These include ***purposive samples, snowball samples, quota samples, and convenience samples***. While the latter two strategies may be used by quantitative researchers from time to time, they are more typically employed in qualitative research, and because they are both nonprobability methods, we include them in this section of the chapter.

To draw a ***purposive sample***, a researcher begins with specific perspectives in mind that he or she wishes to examine and then seeks out research participants who cover that full range of perspectives. For example, if you are studying students' satisfaction with their living quarters on campus, you'll want to be sure to include students who stay in each of the different types or locations of on-campus housing in your study. If you only include students from 1 of 10 dorms on campus, you may miss important details about the experiences of students who live in the 9 dorms you didn't include in your study. Research with young people concerning their workplace sexual harassment experiences, it would be appropriate to choose a purposive sampling strategy. Using participants' prior responses on a survey can ensure that one includes both men and women who'd had a range of harassment experiences in the interviews.

While purposive sampling is often used when one's goal is to include participants who represent a broad range of perspectives, purposive sampling may also be used when a researcher wishes to include only people who meet very narrow or specific criteria. For example, in their study of Japanese women's perceptions of intimate partner violence, Miyoko Nagae and Barbara L. Dancy (2010) [2] limited their study only to participants who had experienced intimate partner violence themselves, were at least 18 years old, had been married and living with their spouse at the time that the violence occurred, were heterosexual, and were willing to be interviewed. In this case, the researchers' goal was to find participants who had had very specific experiences rather than finding those who had had quite diverse experiences, as in the preceding example. In both cases, the researchers involved shared the goal of understanding the topic at hand in as much depth as possible.

Qualitative—and occasionally quantitative—researchers sometimes rely on ***snowball sampling*** techniques to identify study participants. In this case, a researcher might know of one or two people she'd like to include in her study but then relies on those initial participants to help identify additional study participants. Thus, the researcher's sample builds and becomes larger as the study continues, much as a snowball builds and becomes larger as it rolls through the snow.

Snowball sampling is an especially useful strategy when a researcher wishes to study some stigmatized group or behavior. For example, a researcher who wanted to study how people with genital herpes cope with their medical condition would be unlikely to find many participants by posting a call for interviewees in the newspaper or making an announcement about the study at some large social gathering. Instead, the researcher might know someone with the condition, interview that person, and then be referred by the first interviewee to another potential subject. Having a previous participant vouch for the trustworthiness of the researcher may help new potential participants feel more comfortable about being included in the study.

Snowball sampling is sometimes referred to as chain referral sampling. One research participant refers another, and that person refers another, and that person refers another—thus a chain of potential participants is identified. In addition to using this sampling strategy for potentially stigmatized populations, it is also a useful strategy to use when the researcher's group of interest is likely to be difficult to find, not only because of some stigma associated with the group, but also because the group may be relatively rare. This was the case for Steven

M. Kogan and colleagues (Kogan, Wejnert, Chen, Brody, & Slater, 2011) who wished to study the sexual behaviors of non-university-bound African American young adults who lived in high-poverty rural areas. The researchers first relied on their own networks to identify study participants, but because members of the study's target population were not easy to find, access to the networks of initial study participants was very important for identifying additional participants. Initial participants were given coupons to pass on to others they knew who qualified for the study. Participants were given an added incentive for referring eligible study participants; they received not only $50.00 for participating in the study but also $20.00 for each person they recruited who also participated in the study. Using this strategy, Kogan and colleagues succeeded in recruiting 292 study participants.

Quota sampling is another nonprobability sampling strategy. Both qualitative and quantitative researchers regularly employ this type of sampling. When conducting quota sampling, a researcher identifies categories that are important to the study and for which there is likely to be some variation. Subgroups are created based on each category and the researcher decides how many people (or documents or whatever element happens to be the focus of the research) to include from each subgroup and collects data from that number for each subgroup.

Let's go back to the example we considered previously of student satisfaction with on-campus housing. Perhaps there are two types of housing on your campus: apartments that include full kitchens and dorm rooms where residents do not cook for themselves but eat in a dorm cafeteria. As a researcher, you might wish to understand how satisfaction varies across these two types of

housing arrangements. Perhaps you have the time and resources to interview 20 campus residents, so you decide to interview 10 from each housing type. It is possible as well that your review of literature on the topic suggests that campus housing experiences vary by gender. If that is that case, perhaps you'll decide on four important subgroups: men who live in apartments, women who live in apartments, men who live in dorm rooms, and women who live in dorm rooms. Your quota sample would include five people from each subgroup.

In 1936, up-and-coming pollster George Gallup made history when he successfully predicted the outcome of the presidential election using quota sampling methods. The leading polling entity at the time, The Literary Digest, predicted that Alfred Landon would beat Franklin Roosevelt in the presidential election by a landslide. When Gallup's prediction that Roosevelt would win, turned out to be correct, "the Gallup Poll was suddenly on the map" (Van Allen, 2011). Gallup successfully predicted subsequent elections based on quota samples, but in 1948, Gallup incorrectly predicted that Dewey would beat Truman in the US presidential election. Among other problems, the fact that Gallup's quota categories did not represent those who actually voted (Neuman, 2007) underscores the point that one should avoid attempting to make statistical generalizations from data collected using quota sampling methods. While quota sampling offers the strength of helping the researcher account for potentially relevant variation across study elements, it would be a mistake to think of this strategy as yielding statistically representative findings.

Finally, ***convenience sampling*** is another nonprobability sampling strategy that is employed by both qualitative and quantitative researchers. To draw a convenience sample, a researcher simply collects data from those people or other relevant elements to which he or she has most convenient access. This method, also sometimes referred to as haphazard sampling, is most useful in exploratory research. Journalists who need quick and easy access to people from their population of interest also often use it. If you've ever seen brief interviews of people on the street on the news, you've probably seen a haphazard sample being interviewed. While convenience samples offer one major benefit—convenience—we should be cautious about generalizing from research that relies on convenience samples.

Table 6.1: Types of Non-probability Samples

Sample type	Description
Purposive	Researcher seeks out elements that meet specific criteria.
Snowball	Researcher relies on participant referrals to recruit new participants.
Quota	Researcher selects cases from within several different subgroups.
Convenience	Researcher gathers data from whatever cases happen to be convenient.

Sampling for Generalizability

Quantitative researchers are often interested in being able to make generalizations about groups larger than their study samples. While there are certainly instances when quantitative researchers rely on nonprobability samples (e.g., when doing exploratory or evaluation research), quantitative researchers tend to rely on probability sampling techniques. The goals and techniques associated with probability samples differ from those of nonprobability samples. We'll explore those unique goals and techniques in this section.

Probability Sampling

Unlike nonprobability sampling, ***probability sampling*** refers to sampling techniques for which a person's (or event's) likelihood of being selected for membership in the sample is known. You might ask yourself why we should care about a study element's likelihood of being selected for membership in a researcher's sample. The reason is that, in most cases, researchers who use probability sampling techniques are aiming to identify a representative sample from which to collect data. A ***representative sample*** is one that resembles the population from which it was drawn in all the ways that are important for the research being conducted. If, for example, you wish to be able to say something about differences between men and women at the end of your study, you better make sure that your sample doesn't contain only women. That's a bit of an oversimplification, but the point with representativeness is that if your population varies in some way that is important to your study, your sample should contain the same sorts of variation.

Obtaining a representative sample is important in probability sampling because a key goal of studies that rely on probability samples is generalizability. In fact, generalizability is perhaps the key feature that distinguishes probability samples from nonprobability samples. Generalizability refers to the idea that a study's results will tell us something about a group larger than the sample from which the findings were generated. In order to achieve generalizability, a core principle of probability sampling is that all elements in the researcher's target population have an equal chance of being selected for inclusion in the study. In research, this is the principle of random selection. Random selection is a mathematical process that we won't go into too much depth about here, but if you have taken or plan to take a statistics course, you'll learn more about it there. The important thing to remember about random selection here is that, as previously noted, it is a core principal of probability sampling. If a researcher uses random selection techniques to draw a sample, he or she

will be able to estimate how closely the sample represents the larger population from which it was drawn by estimating the sampling error. Sampling error is a statistical calculation of the difference between results from a sample and the actual parameters of a population.

Types of Probability Samples

There are a variety of probability samples that researchers may use. These include ***simple random samples, systematic samples, stratified samples, and cluster samples***.

Simple random samples are the most basic type of probability sample, but their use is not particularly common. Part of the reason for this may be the work involved in generating a simple random sample. To draw a simple random sample, a researcher starts with a list of every single member, or element, of his or her population of interest. This list is sometimes referred to as a ***sampling frame***. Once that list has been created, the researcher numbers each element sequentially and then randomly selects the elements from which he or she will collect data. To randomly select elements, researchers use a table of numbers that have been generated randomly. There are several possible sources for obtaining a random number table. Some statistics and research methods textbooks offer such tables as appendices to the text. Perhaps a more accessible source is one of the many free random number generators available on the Internet. A good online source is the website Stat Trek, which contains a random number generator that you can use to create a random number table of whatever size you might need (https://stattrek.com/Tables/Random.aspx). Randomizer.org also offers a useful random number generator (https://randomizer.org).

As you might have guessed, drawing a simple random sample can be quite tedious. ***Systematic sampling*** techniques are somewhat less tedious but offer the benefits of a random sample. As with simple random samples, you must be able to produce a list of every one of your population elements. Once you've done that, to draw a systematic sample you'd simply select every kth element on your list. But what is k, and where on the list of population elements does one begin the selection process? k is your selection interval or the distance between the elements you select for inclusion in your study. To begin the selection process, you'll need to figure out how many elements you wish to include in your sample. Let's say you want to interview 25 fraternity members on your campus, and there are 100 men on campus who are members of fraternities. In this case, your selection interval, or k, is 4. To arrive at 4, simply divide the total number of population elements by your desired sample size.

To determine where on your list of population elements to begin selecting the names of the 25 men you will interview, select a random number between 1 and k, and begin there. If we randomly select 3 as our starting point, we'd begin by selecting the third fraternity member on the list and then select every fourth member from there. This might be easier to understand if you can see it

visually. lists the names of our hypothetical 100 fraternity members on campus. You'll see that the third name on the list has been selected for inclusion in our hypothetical study, as has every fourth name after that. A total of 25 names have been selected.

There is one clear instance in which systematic sampling should not be employed. If your sampling frame has any pattern to it, you could inadvertently introduce bias into your sample by using a systemic sampling strategy. This is sometimes referred to as the problem of periodicity. Periodicity refers to the tendency for a pattern to occur at regular intervals. Let's say, for example, that you wanted to observe how people use the outdoor public spaces on your campus. Perhaps you need to have your observations completed within 28 days and you wish to conduct four observations on randomly chosen days. Table 6.2 shows a list of the population elements for this example. To determine which days we'll conduct our observations, we'll need to determine our selection interval. As you'll recall from the preceding paragraphs, to do so we must divide our population size, in this case 28 days, by our desired sample size, in this case 4 days. This formula leads us to a selection interval of 7. If we randomly select 2 as our starting point and select every seventh day after that, we'll wind up with a total of 4 days on which to conduct our observations. You'll see how that works out in the following table.

Table 6.2 Systematic Sample of Observation Days

Number	Day	Include in study?	Number	Day	Include in study?
1	Monday		15	Monday	
2	Tuesday	Yes	16	Tuesday	Yes
3	Wednesday		17	Wednesday	
4	Thursday		18	Thursday	
5	Friday		19	Friday	
6	Saturday		20	Saturday	
7	Sunday		21	Sunday	
8	Monday		22	Monday	
9	Tuesday	Yes	23	Tuesday	Yes
10	Wednesday		24	Wednesday	
11	Thursday		25	Thursday	
12	Friday		26	Friday	
13	Saturday		27	Saturday	
14	Sunday		28	Sunday	

Do you notice any problems with our selection of observation days? Apparently we'll only be observing on Tuesdays. As you have probably figured out, that isn't such a good plan if we really wish to understand how public spaces on campus are used. My guess is that weekend use probably differs from weekday use, and that use may even vary during the week, just as class schedules do. In cases such as this, where the sampling frame is cyclical, it would be better to use a stratified sampling technique. In stratified sampling, a researcher will divide the study population into relevant subgroups and then draw a sample from each subgroup. In this example, we might wish to first divide our sampling frame into two lists: weekend days and weekdays. Once we have our two lists, we can then apply either simple random or systematic sampling techniques to each subgroup.

Stratified sampling is a good technique to use when, as in our example, a subgroup of interest makes up a relatively small proportion of the overall sample. In our example of a study of use of public space on campus, we want to be sure to include weekdays and weekends in our sample, but because weekends make up less than a third of an entire week, there's a chance that a simple random or systematic strategy would not yield sufficient weekend observation days. As you might imagine, stratified sampling is even more useful in cases where a subgroup makes up an even smaller proportion of the study population, say, for example, if we want to be sure to include both men's and women's perspectives in a study, but men make up only a small percentage of the population. There's a chance simple random or systematic sampling strategy might not yield any male participants, but by using stratified sampling, we could ensure that our sample contained the proportion of men that is reflective of the larger population.

Up to this point in our discussion of probability samples, we've assumed that researchers will be able to access a list of population elements in order to create a sampling frame. This, as you might imagine, is not always the case. Let's say, for example, that you wish to conduct a study of hairstyle preferences across the United States. Just imagine trying to create a list of every single person with (and without) hair in the country. Basically, we're talking about a list of every person in the country. Even if you could find a way to generate such a list, attempting to do so might not be the most practical use of your time or resources. When this is the case, researchers turn to cluster sampling. ***Cluster sampling*** occurs when a researcher begins by sampling groups (or clusters) of population elements and then selects elements from within those groups.

Let's take a look at a couple more examples. Perhaps you are interested in the workplace experiences of public librarians. Chances are good that obtaining a list of all librarians that work for public libraries would be rather difficult. But I'll bet you could come up with a list of all public libraries without too much hassle. Thus you could draw a random sample of libraries (your cluster) and then draw another random sample of elements (in this case, librarians) from within the libraries you initially selected. Cluster sampling works in stages. In this example, we sampled in two stages. As you might have guessed, sampling in multiple stages does introduce the possibility of greater error (each stage is subject to its own sampling error), but it is nevertheless a highly efficient method.

Holt and Gillespie (2008) used cluster sampling in their study of students' experiences with violence in intimate relationships. Specifically, the researchers randomly selected 14 classes on their campus and then drew a random subsample of students from those classes. But you probably know from your experience with university classes that not all classes are the same size. So if Holt and Gillespie had simply randomly selected 14 classes and then selected the same number of students from each class to complete their survey, then students in the smaller of those classes would have had a greater chance of being selected for the study than students in the larger classes. Keep in mind with random sampling the goal is to make sure that each element has the same chance of being selected. When clusters are of different sizes, as in the example of sampling university classes, researchers often use a method called probability proportionate to size (PPS). This means

that they take into account that their clusters are of different sizes. They do this by giving clusters different chances of being selected based on their size so that each element within those clusters winds up having an equal chance of being selected.

Table 6.3: Types of Probability Samples

Sample type	Description
Simple random	Researcher randomly selects elements from sampling frame.
Systematic	Researcher selects every kth element from sampling frame.
Stratified	Researcher creates subgroups then randomly selects elements from each subgroup.
Cluster	Researcher randomly selects clusters then randomly selects elements from selected clusters.

A Word of Caution: Questions to Ask About Samples

We read and hear about research results so often that we might overlook the need to ask important questions about where research participants come from and how they are identified for inclusion in a research project. It is easy to focus only on findings when we're busy and when the really interesting stuff is in a study's conclusions, not its procedures. But now that you have some familiarity with the variety of procedures for selecting study participants, you are equipped to ask some very important questions about the findings you read and to be a more responsible consumer of research.

Who Sampled, How Sampled, and for What Purpose?

Have you ever been a participant in someone's research? If you have ever taken an introductory psychology or sociology class at a large university, that's probably a silly question to ask. Social science researchers on university campuses have a luxury that researchers elsewhere may not share—they have access to a whole bunch of (presumably) willing and able human guinea pigs. But that luxury comes at a cost—sample representativeness. One study of top academic journals in psychology found that over two-thirds (68%) of participants in studies published by those journals were based on samples drawn in the United States (Arnett, 2008). [1] Further, the study found that two-thirds of the work that derived from US samples published in the Journal of Personality and Social Psychology was based on samples made up entirely of American undergraduates taking psychology courses.

These findings certainly beg the question: What do we actually learn from social scientific studies and about whom do we learn it? That is exactly the concern raised by Henrich and colleagues (Henrich, Heine, & Norenzayan, 2010), authors of the article "The Weirdest People in the World?" In their piece, Henrich and colleagues point out that behavioral scientists very commonly make sweeping claims about human nature based on samples drawn only from WEIRD (Western, educated, industrialized, rich, and democratic) societies, and often based on even narrower samples, as is the case with many studies relying on samples drawn from university classrooms. As it turns out, many robust findings about the nature of human behavior when it comes to fairness, cooperation, visual perception, trust, and other behaviors are based on studies that excluded participants from outside the United States and sometimes excluded anyone outside the university classroom (Begley, 2010). This certainly raises questions about what we really know about human behavior

as opposed to US resident or US undergraduate behavior. Of course not all research findings are based on samples of WEIRD folks like university students. But even then it would behoove us to pay attention to the population on which studies are based and the claims that are being made about to whom those studies apply.

In the preceding discussion, the concern is with researchers making claims about populations other than those from which their samples were drawn. A related, but slightly different, potential concern is sampling bias. Bias in sampling occurs when the elements selected for inclusion in a study do not represent the larger population from which they were drawn. For example, a poll conducted online by a newspaper asking for the public's opinion about some local issue will certainly not represent the public since those without access to computers or the Internet, those who do not read that paper's website, and those who do not have the time or interest will not answer the question.

Another thing to keep in mind is that just because a sample may be representative in all respects that a researcher thinks are relevant, there may be aspects that are relevant that didn't occur to the researcher when she was drawing her sample. You might not think that a person's phone would have much to do with their voting preferences, for example. But had pollsters making predictions about the results of the 2008 presidential election not been careful to include both cell phone–only and landline households in their surveys, it is possible that their predictions would have underestimated Barack Obama's lead over John McCain because Obama was much more popular among cell-only users than McCain (Keeter, Dimock, & Christian, 2008).

So how do we know when we can count on results that are being reported to us? While there might not be any magic or always-true rules we can apply, there are a couple of things we can keep in mind as we read the claims researchers make about their findings. First, remember that sample quality is determined only by the sample actually obtained, not by the sampling method itself. A researcher may set out to administer a survey to a representative sample by correctly employing a random selection technique, but if only a handful of the people sampled actually respond to the survey, the researcher will have to be very careful about the claims he can make about his survey findings. Another thing to keep in mind, as demonstrated by the preceding discussion, is that researchers may be drawn to talking about implications of their findings as though they apply to some group other than the population actually sampled. Though this tendency is usually quite innocent and does not come from a place of malice, it is all too tempting a way to talk about findings; as consumers of those findings, it is our responsibility to be attentive to this sort of (likely unintentional) bait and switch.

Finally, keep in mind that a sample that allows for comparisons of theoretically important concepts or variables is certainly better than one that does not allow for such comparisons. In a study based on a nonrepresentative sample, for example, we can learn about the strength of our social theories by comparing relevant aspects of social processes. Klawiter's previously mentioned study (1999) [5] of three carefully chosen breast cancer activist groups allowed her to contribute to our understandings of activism by addressing potential weaknesses in theories of social change.

At their core, questions about sample quality should address who has been sampled, how they were sampled, and for what purpose they were sampled. Being able to answer those questions will help you better understand, and more responsibly read, research results.

Key Takeaways, Exercises, and References

Key Takeaways

- A population is the group that is the main focus of a researcher's interest; a sample is the group from whom the researcher actually collects data.
- Populations and samples might be one and the same, but more often they are not.
- Sampling involves selecting the observations that you will analyze.
- Nonprobability samples might be used when researchers are conducting exploratory research, by evaluation researchers, or by researchers whose aim is to make some theoretical contribution.
- There are several types of nonprobability samples including purposive samples, snowball samples, quota samples, and convenience samples.
- Sometimes researchers may make claims about populations other than those from whom their samples were drawn; other times they may make claims about a population based on a sample that is not representative. As consumers of research, we should be attentive to both possibilities.
- A researcher's findings need not be generalizable to be valuable; samples that allow for comparisons of theoretically important concepts or variables may yield findings that contribute to our social theories and our understandings of social processes.
- In probability sampling, the aim is to identify a sample that resembles the population from which it was drawn.
- There are several types of probability samples including simple random samples, systematic samples, stratified samples, and cluster samples.

Exercises

- Read through the methods section of a couple of scholarly articles describing empirical research. How do the authors talk about their populations and samples, if at all? What do the articles' abstracts suggest in terms of whom conclusions are being drawn about?
- Think of a research project you have envisioned conducting as you've read this text. Would your population and sample be one and the same, or would they differ somehow? Explain.
- Imagine you are about to conduct a study of people's use of the public parks in your home-

town. Explain how you could employ each of the nonprobability sampling techniques described previously to recruit a sample for your study.

- Of the four nonprobability sample types described, which seems strongest to you? Which seems weakest? Explain.
- Find any news story or blog entry that describes results from any social scientific study. How much detail is reported about the study's sample? What sorts of claims are being made about the study's findings, and to whom do they apply?
- Imagine that you are about to conduct a study of people's use of public parks. Explain how you could employ each of the probability sampling techniques described earlier to recruit a sample for your study.
- Of the four probability sample types described, which seems strongest to you? Which seems weakest? Explain.

References

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. American Psychologist, 63, 602–614.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33, 61–135.

Holt, J. L., & Gillespie, W. (2008). Intergenerational transmission of violence, threatened egoism, and reciprocity: A test of multiple pychosocial factors affecting intimate partner violence.American Journal of Criminal Justice, 33, 252–266.

Keeter, S., Dimock, M., & Christian, L. (2008). Calling cell phones in '08 pre-election polls. The Pew Research Center for the People and the Press. Retrieved from https://people-press.org/ https://people-press.org/files/legacy-pdf/cell-phone- commentary.pdf

Klawiter, M. (1999). Racing for the cure, walking women, and toxic touring: Mapping cultures of action within the Bay Area terrain of breast cancer. Social Problems, 46, 104–126.

Klawiter, M. (1999). Racing for the cure, walking women, and toxic touring: Mapping cultures of action within the Bay Area terrain of breast cancer. Social Problems, 46, 104–126.

Kogan, S. M., Wejnert, C., Chen, Y., Brody, G. H., & Slater, L. M. (2011). Respondent-driven sampling with hard-to-reach emerging adults: An introduction and case study with rural African Americans. Journal of Adolescent Research, 26, 30–60.

Nagae, M., & Dancy, B. L. (2010). Japanese women's perceptions of intimate partner violence (IPV). Journal of Interpersonal Violence, 25, 753– 766.

Neuman, W. L. (2007). Basics of social research: Qualitative and quantitative approaches (2nd ed.). Boston, MA: Pearson.

Newsweek magazine published an interesting story about Henrich and his colleague's study: Begley, S. (2010). What's really human? The trouble with student guinea pigs. Retrieved from https://www.newsweek.com/2010/07/23/what-s-really- human.html

Van Allen, S. (2011). Gallup corporate history. Retrieved from https://www.gallup.com/corporate/1357/Corporate- History.aspx#2

CHAPTER 7: SURVEY RESEARCH

In 2008, the voters of the United States elected their first African American president, Barack Obama. It may not surprise you to learn that when President Obama was coming of age in the 1970s, one-quarter of Americans reported that they would not vote for a qualified African American presidential nominee. Three decades later, when President Obama ran for the presidency, fewer than 8% of Americans still held that position, and President Obama won the election (Smith, 2009). We know about these trends in voter opinion because the General Social Survey (https://www.norc.uchicago.edu/GSS+Website), a nationally representative survey of American adults, included questions about race and voting over the years described here. Without survey research, we may not know how Americans' perspectives on race and the presidency shifted over these years.

LEARNING OBJECTIVES

- Define survey research.
- Identify when it is appropriate to employ survey research as a data-collection strategy.
- Identify and explain the strengths of survey research.
- Identify and explain the weaknesses of survey research.
- Define response rate, and discuss some of the current thinking about response rates.
- Define cross-sectional surveys, provide an example of a cross-sectional survey, and outline some of the drawbacks of cross-sectional research.
- Describe the various types of longitudinal surveys.
- Define retrospective surveys, and identify their strengths and weaknesses.
- Discuss some of the benefits and drawbacks of the various methods of delivering self-administered questionnaires.
- Identify the steps one should take in order to write effective survey questions.
- Describe some of the ways that survey questions might confuse respondents and how to overcome that possibility.
- Recite the two response option guidelines when writing closed-ended questions.
- Define fence-sitting and floating.
- Describe the steps involved in constructing a well-designed questionnaire.
- Discuss why pretesting is important.

Survey Research: What Is It and When Should It Be Used?

Most of you have probably taken a survey at one time or another, so you probably have a pretty good idea of what a survey is. Sometimes students in my research methods classes feel that understanding what a survey is and how to write one is so obvious, there's no need to dedicate any class time to learning about it. This feeling is understandable—surveys are very much a part of our everyday lives—we've probably all taken one, we hear about their results in the news, and perhaps we've even administered one ourselves. What students quickly learn is that there is more to constructing a good survey than meets the eye. Survey design takes a great deal of thoughtful planning and often a great many rounds of revision, but it is worth the effort. As we'll learn in this chapter, there are many benefits to choosing survey research as one's method of data collection. We'll take a look at what a survey is exactly, what some of the benefits and drawbacks of this method are, how to construct a survey, and what to do with survey data once one has it in hand.

Survey research is a quantitative method whereby a researcher poses some set of predetermined questions to an entire group, or sample, of individuals. Survey research is an especially useful approach when a researcher aims to describe or explain features of a very large group or groups. This method may also be used as a way of quickly gaining some general details about one's population of interest to help prepare for a more focused, in-depth study using time-intensive methods such as in-depth interviews or field research. In this case, a survey may help a researcher identify specific individuals or locations from which to collect additional data.

As is true of all methods of data collection, survey research is better suited to answering some kinds of research question more than others. In addition, as you'll recall, operationalization works differently with different research methods. If your interest is in political activism, for example, you likely operationalize that concept differently in a survey than you would for a field research study of the same topic.

Pros and Cons of Survey Research

Strengths of Survey Method

Researchers employing survey methods to collect data enjoy a number of benefits. First, surveys are an excellent way to gather lots of information from many people. In my own study of older people's experiences in the workplace, I was able to mail a written questionnaire to around 500 people who lived throughout the state of Maine at a cost of just over $1,000. This cost included printing copies of my seven-page survey, printing a cover letter, addressing and stuffing envelopes, mailing the survey, and buying return postage for the survey. I realize that $1,000 is nothing to sneeze at. But just imagine what it might have cost to visit each of those people individually to interview them in person. Consider the cost of gas to drive around the state, other travel costs, such as meals and lodging while on the road, and the cost of time to drive to and talk with each person individually. We could double, triple, or even quadruple our costs pretty quickly by opting for an in-person method of data collection over a mailed survey. Thus, surveys are relatively cost effective.

Related to the benefit of cost effectiveness is a survey's potential for generalizability. Because surveys allow researchers to collect data from very large samples for a relatively low cost, survey methods lend themselves to probability sampling techniques. Of all the data-collection methods described in this text, survey research is probably the best method to use when one hopes to gain a representative picture of the attitudes and characteristics of a large group.

Survey research also tends to be a reliable method of inquiry. This is because surveys are standardized in that the same questions, phrased in exactly the same way, are posed to participants. This is not to say that all surveys are always reliable. A poorly phrased question can cause respondents to interpret its meaning differently, which can reduce that question's reliability. Assuming well-constructed question and questionnaire design, one strength of survey methodology is its potential to produce reliable results.

The versatility of survey research is also an asset. Surveys are used by all kinds of people in all kinds of professions. I repeat, surveys are used by all kinds of people in all kinds of professions. Is there a light bulb switching on in your head? I hope so. The versatility offered by survey research means that understanding how to construct and administer surveys is a useful skill to have for all kinds of jobs. For example, social service and other organizations use surveys to evaluate the effectiveness of their efforts, businesses use them to learn how to market their products, governments use them to understand community opinions and needs, and politicians and media outlets use surveys to understand their constituencies.

In sum, the following are benefits of survey research:

- Cost-effective
- Generalizable
- Reliable
- Versatile

Weaknesses of Survey Method

As with all methods of data collection, survey research also comes with a few drawbacks. First, while one might argue that surveys are flexible in the sense that we can ask any number of questions on any number of topics in them, the fact that the survey researcher is generally stuck with a single instrument for collecting data (the questionnaire), surveys are in many ways rather inflexible. Let's say you mail a survey out to 1,000 people and then discover, as responses start coming in, that your phrasing on a particular question seems to be confusing a number of respondents. At this stage, it's too late for a do-over or to change the question for the respondents who haven't yet returned their surveys. When conducting in-depth interviews, on the other hand, a researcher can provide respondents further explanation if they're confused by a question and can tweak their questions as they learn more about how respondents seem to understand them.

Validity can also be a problem with surveys. Survey questions are standardized; thus it can be difficult to ask anything other than very general questions that a broad range of people will understand. Because of this, survey results may not be as valid as results obtained using methods of data collection that allow a researcher to more comprehensively examine whatever topic is being studied. Let's say, for example, that you want to learn something about voters' willingness to elect an African American president, as in our opening example in this chapter. General Social Survey respondents were asked, "If your party nominated an African American for president, would you vote for him if he were qualified for the job?" Respondents were then asked to respond either yes or no to the question. But what if someone's opinion was more complex than could be answered with a simple yes or no? What if, for example, a person was willing to vote for an African American woman but not an African American man?

Types of Surveys

There is much variety when it comes to surveys. This variety comes both in terms of time—when or with what frequency a survey is administered—and in terms of administration—how a survey is delivered to respondents. In this section we'll take a look at what types of surveys exist when it comes to both time and administration.

The Time Dimension

In terms of time, there are two main types of surveys: ***cross-sectional and longitudinal***. Cross-sectional surveys are those that are administered at just one point in time. These surveys offer researchers a sort of snapshot in time and give us an idea about how things are for our respondents at the particular point in time that the survey is administered.

An example of a cross-sectional survey comes from Kezdy and colleagues' study (Kezdy, Martos, Boland, & Horvath-Szabo, 2011) of the association between religious attitudes, religious beliefs, and mental health among students in Hungary. These researchers administered a single, one-time-only, cross-sectional survey to a convenience sample of 403 high school and university students. The survey focused on how religious attitudes impact various aspects of one's life and health. The researchers found from analysis of their cross- sectional data that anxiety and depression were highest among those who had both strong religious beliefs and also some doubts about religion. Yet another example of cross-sectional survey research can be seen in Bateman and colleagues' study (Bateman, Pike, & Butler, 2011) of how the perceived publicness of social networking sites influences users' self-disclosures. These researchers administered an online survey to undergraduate and graduate business students. They found that even though revealing information about oneself is viewed as key to realizing many of the benefits of social networking sites, respondents were less willing to disclose information about themselves as their perceptions of a social networking site's publicness rose. That is, there was a negative relationship between perceived publicness of a social networking site and plans to self- disclose on the site.

One problem with cross-sectional surveys is that the events, opinions, behaviors, and other phenomena that such surveys are designed to assess don't generally remain stagnant. Thus generalizing from a cross-sectional survey about the way things are can be tricky; perhaps you can say something about the way things were in the moment that you administered your survey, but it is difficult to know whether things remained that way for long after you administered your survey. Think, for example, about how Americans might have responded if administered a survey asking for their opinions on terrorism on September 10, 2001. Now imagine how responses to the same set of

questions might differ were they administered on September 12, 2001. The point is not that cross-sectional surveys are useless; they have many important uses. Researchers, though, must remember what they have captured by administering a cross-sectional survey; that is, a snapshot of life as it was at the time that the survey was administered.

One way to overcome this sometimes problematic aspect of cross-sectional surveys is to administer a longitudinal survey. ***Longitudinal surveys*** are those that enable a researcher to make observations over some extended period of time. There are several types of longitudinal surveys, including trend, panel, and cohort surveys. We'll discuss all three types here, along with another type of survey called retrospective. Retrospective surveys fall somewhere in between cross-sectional and longitudinal surveys.

The first type of longitudinal survey is called a ***trend survey***. The main focus of a trend survey is, perhaps not surprisingly, trends. Researchers conducting trend surveys are interested in how people's inclinations change over time. The Gallup opinion polls are an excellent example of trend surveys. You can read more about Gallup on their website: https://www.gallup.com/Home.aspx. To learn about how public opinion changes over time, Gallup administers the same questions to people at different points in time. For example, for several years Gallup has polled Americans to find out what they think about gas prices (something many of us happen to have opinions about). One thing we've learned from Gallup's polling is that price increases in gasoline caused financial hardship for 67% of respondents in 2011, up from 40% in the year 2000. Gallup's findings about trends in opinions about gas prices have also taught us that whereas just 34% of people in early 2000 thought the current rise in gas prices was permanent, 54% of people in 2011 believed the rise to be permanent. Thus through Gallup's use of trend survey methodology, we've learned that Americans seem to feel generally less optimistic about the price of gas these days than they did 10 or so years ago. It should be noted that in a trend survey, the same people are probably not answering the researcher's questions each year. Because the interest here is in trends, not specific people, as long as the researcher's sample is representative of whatever population he or she wishes to describe trends for, it isn't important that the same people participate each time.

A specialized form of the above are ***panel surveys***. Unlike in a trend survey, in a panel survey the same people do participate in the survey each time it is administered. As you might imagine, panel studies can be difficult and costly. Imagine trying to administer a survey to the same 100 people every year for, say, 5 years in a row. Keeping track of where people live, when they move, and when they die takes resources that researchers often don't have. When they do, however, the results can be quite powerful. The Youth Development Study (YDS), administered from the University of Minnesota, offers an excellent example of a panel study. You can read more about the Youth Development Study at its website: https://www.soc.umn.edu/research/yds. Since 1988, YDS researchers have administered an annual survey to the same 1,000 people. Study participants were in ninth grade when the study began, and they are now in their thirties. Several hundred papers, articles, and books have been written using data from the YDS. One of the major lessons learned from this panel study is that work has a largely positive impact on young people (Mortimer, 2003). Contrary

to popular beliefs about the impact of work on adolescents' performance in school and transition to adulthood, working increases confidence, enhances academic success, and prepares students for success in their future careers. Without this panel study, we may not be aware of the positive impact that working can have on young people.

Another type of longitudinal survey is a ***cohort survey***. In a cohort survey, a researcher identifies some category of people that are of interest and then regularly surveys people who fall into that category. The same people don't necessarily participate from year to year, but all participants must meet whatever categorical criteria fulfill the researcher's primary interest. Common cohorts that may be of interest to researchers include people of particular generations or those who were born around the same time period, graduating classes, people who began work in a given industry at the same time, or perhaps people who have some specific life experience in common. An example of this sort of research can be seen in Percheski's work (2008) on cohort differences in women's employment. Percheski compared women's employment rates across seven different generational cohorts, from Progressives born between 1906 and 1915 to Generation Xers born between 1966 and 1975. She found, among other patterns, that professional women's labor force participation had increased across all cohorts. She also found that professional women with young children from Generation X had higher labor force participation rates than similar women from previous generations, concluding that mothers do not appear to be opting out of the workforce as some journalists have speculated (Belkin, 2003).

All three types of longitudinal surveys share the strength that they permit a researcher to make observations over time. This means that if whatever behavior or other phenomenon the researcher is interested in changes, either because of some world event or because people age, the researcher will be able to capture those changes. Table 8.1 "Types of Longitudinal Surveys" summarizes each of the three types of longitudinal surveys.

Table 7.1: Types of Longitudinal Surveys

Sample type	Description
Trend	Researcher examines changes in trends over time; the same people do not necessarily participate in the survey more than once.
Panel	Researcher surveys the exact same sample several times over a period of time.
Cohort	Researcher identifies some category of people that are of interest and then regularly surveys people who fall into that category.

Finally, ***retrospective*** surveys are similar to other longitudinal studies in that they deal with changes over time, but like a cross-sectional study, they are administered only once. In a retrospective survey, participants are asked to report events from the past. By having respondents report past

behaviors, beliefs, or experiences, researchers are able to gather longitudinal-like data without actually incurring the time or expense of a longitudinal survey. Of course, this benefit must be weighed against the possibility that people's recollections of their pasts may be faulty. Imagine, for example, that you're asked in a survey to respond to questions about where, how, and with whom you spent last Valentine's Day. As last Valentine's Day can't have been more than 12 months ago, chances are good that you might be able to respond accurately to any survey questions about it. But now let's say the research wants to know how last Valentine's Day compares to previous Valentine's Days, so he asks you to report on where, how, and with whom you spent the preceding six Valentine's Days. How likely is it that you will remember? Will your responses be as accurate as they might have been had you been asked the question each year over the past 6 years rather than asked to report on all years today?

In sum, when or with what frequency a survey is administered will determine whether your survey is cross- sectional or longitudinal. While longitudinal surveys are certainly preferable in terms of their ability to track changes over time, the time and cost required to administer a longitudinal survey can be prohibitive. As you may have guessed, the issues of time described here are not necessarily unique to survey research. Other methods of data collection can be cross-sectional or longitudinal—these are really matters of research design. But we've placed our discussion of these terms here because they are most commonly used by survey researchers to describe the type of survey administered. Another aspect of survey administration deals with how surveys are administered.

Administration

Surveys vary not just in terms of when they are administered but also in terms of how they are administered. One common way to administer surveys is in the form of self-administered questionnaires. This means that a research participant is given a set of questions, in writing, to which he or she is asked to respond. Self- administered questionnaires can be delivered in hard copy format, typically via mail, or increasingly more commonly, online. We'll consider both modes of delivery here.

Hard copy self-administered questionnaires may be delivered to participants in person or via snail mail. Perhaps you've take a survey that was given to you in person; on many university campuses it is not uncommon for researchers to administer surveys in large social science classes. In my own courses, I've welcomed graduate students and professors doing research in areas that are relevant to my students, such as studies of campus life, to administer their surveys to the class. If you are ever asked to complete a survey in a similar setting, it might be interesting to note how your perspective on the survey and its questions could be shaped by the new knowledge you're gaining about survey research in this chapter.

Researchers may also deliver surveys in person by going door-to-door and either asking people to fill them out right away or making arrangements for the researcher to return to pick up completed surveys. Though the advent of online survey tools has made door-to-door delivery of surveys less common, I still see an occasional survey researcher at my door, especially around election time. This mode of gathering data is apparently still used by political campaign workers, at least in some areas of the country.

If you are not able to visit each member of your sample personally to deliver a survey, you might consider sending your survey through the mail. While this mode of delivery may not be ideal (imagine how much less likely you'd probably be to return a survey that didn't come with the researcher standing on your doorstep waiting to take it from you), sometimes it is the only available or the most practical option. As I've said, this may not be the most ideal way of administering a survey because it can be difficult to convince people to take the time to complete and return your survey.

Often survey researchers who deliver their surveys via snail mail may provide some advance notice to respondents about the survey to get people thinking about and preparing to complete it. They may also follow up with their sample a few weeks after their survey has been sent out. This can be done not only to remind those who have not yet completed the survey to please do so but also to thank those who have already returned the survey. Most survey researchers agree that this sort of follow-up is essential for improving mailed surveys' return rates (Babbie, 2010).

Online delivery as another way to administer a survey. This delivery mechanism is becoming increasingly common, no doubt because it is easy to use, relatively cheap, and may be quicker than knocking on doors or waiting for mailed surveys to be returned. To deliver a survey online, a researcher may subscribe to a service that offers online delivery or use some delivery mechanism that is available for free. SurveyMonkey offers both free and paid online survey services (https://www.surveymonkey.com). One advantage to using a service like SurveyMonkey, aside from the advantages of online delivery already mentioned, is that results can be provided to you in formats that are readable by data analysis programs such as SPSS, Systat, and Excel. This saves you the step of having to manually enter data into your analysis program, as you would if you administered your survey in hard copy format.

Many of the suggestions provided for improving the response rate on a hard copy questionnaire apply to online questionnaires as well. One difference, of course, is that the sort of incentives one can provide in an online format differ from those that can be given in person or sent through the mail. But this doesn't mean that online survey researchers cannot offer completion incentives to their respondents. I've taken a number of online surveys; on one, I was given a printable $5 coupon to my university's campus dining services on completion, and another time I was given a coupon code to use for $10 off any order on Amazon.com. I've taken other online surveys where on completion I provided my name and contact information to be entered into a drawing together to win a larger gift.

Sometimes surveys are administered by having a researcher actually pose questions directly to respondents rather than having respondents read the questions on their own. These types of surveys are a form of interviews. Qualitative interview methodology, however, differs from survey research in that data are collected via a personal interaction that is largely conversational. Orally administered surveys are standardized and do not vary from respondent to respondent. The main benefit, and only exception to self-administered surveys, is that a researcher can offer clarifications if questions should arise.

Whatever delivery mechanism you choose, keep in mind that there are pros and cons to each of the options described here. While online surveys may be faster and cheaper than mailed surveys, can you be certain that every person in your sample will have the necessary computer hardware, software, and Internet access in order to complete your online survey? On the other hand, perhaps mailed surveys are more likely to reach your entire sample but also more likely to be lost and not returned. The choice of which delivery mechanism is best depends on a number of factors including your resources, the resources of your study participants, and the time you have available to distribute surveys and wait for responses. In my own survey of older workers, I would have much preferred to administer my survey online, but because so few people in my sample were likely to have computers, and even fewer would have Internet access, I chose instead to mail paper copies of the survey to respondents' homes. Understanding the characteristics of your study's population is key to identifying the appropriate mechanism for delivering your survey.

Designing Effective Questions and Questionnaires

To this point we've considered several general points about surveys including when to use them, some of their pros and cons, and how often and in what ways to administer surveys. In this section we'll get more specific and take a look at how to pose understandable questions that will yield useable data and how to present those questions on your questionnaire.

Asking Effective Questions

The first thing you need to do in order to write effective survey questions is identify what exactly it is that you wish to know. As silly as it sounds to state what seems so completely obvious, I can't stress enough how easy it is to forget to include important questions when designing a survey. Let's say you want to understand how students at your school made the transition from high school to university. Perhaps you wish to identify which students were comparatively more or less successful in this transition and which factors contributed to students' success or lack thereof. To understand which factors shaped successful students' transitions to university, you'll need to include questions in your survey about all the possible factors that could contribute. Consulting the literature on the topic will certainly help, but you should also take the time to do some brainstorming on your own and to talk with others about what they think may be important in the transition to university. Perhaps time or space limitations won't allow you to include every single item you've come up with, so you'll also need to think about ranking your questions so that you can be sure to include those that you view as most important.

Although I have stressed the importance of including questions on all topics you view as important to your overall research question, you don't want to take an everything-but-the-kitchen-sink approach by uncritically including every possible question that occurs to you. Doing so puts an unnecessary burden on your survey respondents. Remember that you have asked your respondents to give you their time and attention and to take care in responding to your questions; show them your respect by only asking questions that you view as important.

Once you've identified all the topics about which you'd like to ask questions, you'll need to actually write those questions. Questions should be as clear and to the point as possible. This is not the time to show off your creative writing skills; a survey is a technical instrument and should be written in a way that is as direct and succinct as possible. As I've said, your survey respondents have

agreed to give their time and attention to your survey. The best way to show your appreciation for their time is to not waste it. Ensuring that your questions are clear and not overly wordy will go a long way toward showing your respondents the gratitude they deserve.

Related to the point about not wasting respondents' time, make sure that every question you pose will be relevant to every person you ask to complete it. This means two things: first, that respondents have knowledge about whatever topic you are asking them about, and second, that respondents have experience with whatever events, behaviors, or feelings you are asking them to report. You probably wouldn't want to ask a sample of 18-year-old respondents, for example, how they would have advised President Reagan to proceed when news of the United States' sale of weapons to Iran broke in the mid-1980s. For one thing, few 18-year-olds are likely to have any clue about how to advise a president (nor does this 30-something- year-old). Furthermore, the 18-year-olds of today were not even alive during Reagan's presidency, so they have had no experience with the event about which they are being questioned. In our example of the transition to university, heeding the criterion of relevance would mean that respondents must understand what exactly you mean by "transition to university" if you are going to use that phrase in your survey and that respondents must have actually experienced the transition to university themselves.

If you decide that you do wish to pose some questions about matters with which only a portion of respondents will have had experience, it may be appropriate to introduce a filter question into your survey. A filter question is designed to identify some subset of survey respondents who are asked additional questions that are not relevant to the entire sample. Perhaps in your survey on the transition to university you want to know whether substance use plays any role in students' transitions. You may ask students how often they drank during their first semester of university. But this assumes that all students drank. Certainly some may have abstained, and it wouldn't make any sense to ask the non-drinkers how often they drank. Nevertheless, it seems reasonable that drinking frequency may have an impact on someone's transition to university, so it is probably worth asking this question even if doing so violates the rule of relevance for some respondents. This is just the sort of instance when a filter question would be appropriate.

There are some ways of asking questions that are bound to confuse a good many survey respondents. Survey researchers should take great care to avoid these kinds of questions. These include questions that pose double negatives, those that use confusing or culturally specific terms, and those that ask more than one question but are posed as a single question. Any time respondents are forced to decipher questions that utilize two forms of negation, confusion is bound to ensue. Taking the previous question about drinking as our example, what if we had instead asked, "Did you not drink during your first semester of university?" A response of no would mean that the respondent did actually drink—he or she did not not drink. This example is obvious, but hopefully it drives home the point to be careful about question wording so that respondents are not asked to decipher double negatives. In general, avoiding negative terms in your question wording will help to increase respondent understanding.

You should also avoid using terms or phrases that may be regionally or culturally specific (unless you are absolutely certain all your respondents come from the region or culture whose terms you are using). When I first moved to Maine from Minnesota, I was totally confused every time I heard someone use the word wicked. This term has totally different meanings across different regions of the country. I'd come from an area that understood the term wicked to be associated with evil. In my new home, however, wicked is used simply to put emphasis on whatever it is that you're talking about. So if this chapter is extremely interesting to you, if you live in Maine you might say that it is "wicked interesting." If you hate this chapter and you live in Minnesota, perhaps you'd describe the chapter simply as wicked. I once overheard one student tell another that his new girlfriend was "wicked athletic." At the time I thought this meant he'd found a woman who used her athleticism for evil purposes. I've come to understand, however, that this woman is probably just exceptionally athletic. While wicked may not be a term you're likely to use in a survey, the point is to be thoughtful and cautious about whatever terminology you do use.

Asking multiple questions as though they are a single question can also be terribly confusing for survey respondents. There's a specific term for this sort of question; it is called a double-barreled question. Using our example of the transition to University, consider: "Compared to your previous schooling, do you find University to be more demanding and interesting?" Do you see what makes the question double-barreled? How would someone respond if they felt their University classes were more demanding but also more boring than their high school classes? Or less demanding but more interesting? Because the question combines "demanding" and "interesting," there is no way to respond yes to one criterion but no to the other.

Another thing to avoid when constructing survey questions is the problem of social desirability. We all want to look good, right? And we all probably know the politically correct response to a variety of questions whether we agree with the politically correct response or not. In survey research, social desirability refers to the idea that respondents will try to answer questions in a way that will present them in a favorable light. Perhaps we decide that to understand the transition to university, we need to know whether respondents ever cheated on an exam in high school or university. We all know that cheating on exams is generally frowned upon (at least I hope we all know this). So, it may be difficult to get people to admit to cheating on a survey. But if you can guarantee respondents' confidentiality, or even better, their anonymity, chances are much better that they will be honest about having engaged in this socially undesirable behavior. Another way to avoid problems of social desirability is to try to phrase difficult questions in the most benign way possible. Earl Babbie (2010) offers a useful suggestion for helping you do this—simply imagine how you would feel responding to your survey questions. If you would be uncomfortable, chances are others would as well.

Finally, it is important to get feedback on your survey questions from as many people as possible, especially people who are like those in your sample. Now is not the time to be shy. Ask your friends for help, ask your mentors for feedback, ask your family to take a look at your survey as well. The

more feedback you can get on your survey questions, the better the chances that you will come up with a set of questions that are understandable to a wide variety of people and, most importantly, to those in your sample.

In sum, in order to pose effective survey questions, researchers should do the following:

- Identify what it is they wish to know.
- Keep questions clear and succinct.
- Make questions relevant to respondents.
- Use filter questions when necessary.
- Avoid questions that are likely to confuse respondents such as those that use double negatives, use culturally specific terms, or pose more than one question in the form of a single question.
- Imagine how they would feel responding to questions.
- Get feedback, especially from people who resemble those in the researcher's sample.

Response Options

While posing clear and understandable questions in your survey is certainly important, so, too, is providing respondents with unambiguous response options. Response options are the answers that you provide to the people taking your survey. Generally respondents will be asked to choose a single (or best) response to each question you pose, though certainly it makes sense in some cases to instruct respondents to choose multiple response options. One caution to keep in mind when accepting multiple responses to a single question, however, is that doing so may add complexity when it comes to tallying and analyzing your survey results.

Offering response options assumes that your questions will be closed-ended questions. In a quantitative written survey, which is the type of survey we've been discussing here, chances are good that most if not all your questions will be closed ended. This means that you, the researcher, will provide respondents with a limited set of options for their responses. To write an effective closed-ended question, there are a couple of guidelines worth following. First, be sure that your response options are mutually exclusive. Look back at Figure 8.8 "Filter Question", which contains questions about how often and how many drinks respondents consumed. Do you notice that there are no overlapping categories in the response options for these questions? This is another one of those points about question construction that seems fairly obvious but that can be easily overlooked. Response options should also be exhaustive. In other words, every possible response should be covered in the set of response options that you provide.

Surveys need not be limited to closed-ended questions. Sometimes survey researchers include open-ended questions in their survey instruments as a way to gather additional details from respondents. An open-ended question does not include response options; instead, respondents are asked to reply to the question in their own way, using their own words. These questions are generally used to find out more about a survey participant's experiences or feelings about whatever they are being asked to report in the survey. If, for example, a survey includes closed-ended questions asking respondents to report on their involvement in extracurricular activities during university, an open-ended question could ask respondents why they participated in those activities or what they gained from their participation. While responses to such questions may also be captured using a closed-ended format, allowing participants to share some of their responses in their own words can make the experience of completing the survey more satisfying to respondents and can also reveal new motivations or explanations that had not occurred to the researcher. However, while these response provide interesting qualitative data, it becomes very difficult to contend with such data quantitatively. In order to do so, all the possible answers that may be provided would need to be able to be slotted into pre-established categories for analysis. For instance, perhaps such open-ended responses to "motivation to participate" might be coded in the categories, "health," "socia-

bility," "change," and "other." As you may well imagine, in order that the "other" category not be overwhelmed, a fair amount of thought needs to be given to the range of possible responses. What other categories of motivation can you think of?

Other things to avoid when it comes to response options include fence-sitting and floating. Fence-sitters are respondents who choose neutral response options, even if they have an opinion. This can occur if respondents are given, say, five rank-ordered response options, such as strongly agree, agree, no opinion, disagree, and strongly disagree. Some people will be drawn to respond "no opinion" even if they have an opinion, particularly if their true opinion is not a socially desirable opinion. Floaters, on the other hand, are those that choose a substantive answer to a question when really they don't understand the question or don't have an opinion. If a respondent is only given four rank-ordered response options, such as strongly agree, agree, disagree, and strongly disagree, those who have no opinion have no choice but to select a response that suggests they have an opinion.

As you can see, floating is the flip side of fence-sitting. Thus the solution to one problem is often the cause of the other. How you decide which approach to take depends on the goals of your research. Sometimes researchers actually want to learn something about people who claim to have no opinion. In this case, allowing for fence- sitting would be necessary. Other times researchers feel confident their respondents will all be familiar with every topic in their survey. In this case, perhaps it is OK to force respondents to choose an opinion. There is no always-correct solution to either problem.

Finally, using a matrix is a nice way of streamlining response options. A matrix is a question type that that lists a set of questions for which the answer categories are all the same. If you have a set of questions for which the response options are the same, it may make sense to create a matrix rather than posing each question and its response options individually. Not only will this save you some space in your survey but it will also help respondents progress through your survey more easily.

Designing Questionnaires

In addition to constructing quality questions and posing clear response options, you'll also need to think about how to present your written questions and response options to survey respondents. Questions are presented on a questionnaire, the document (either hard copy or online) that contains all your survey questions that respondents read and mark their responses on. Designing questionnaires takes some thought, and in this section we'll discuss the sorts of things you should think about as you prepare to present your well-constructed survey questions on a questionnaire.

One of the first things to do once you've come up with a set of survey questions you feel confident about is to group those questions thematically. In our example of the transition to university, perhaps we'd have a few questions asking about study habits, others focused on friendships, and still others on exercise and eating habits. Those may be the themes around which we organize our questions. Or perhaps it would make more sense to present any questions we had about preuniversity life and habits and then present a series of questions about life after beginning university. The point here is to be deliberate about how you present your questions to respondents. Once you have grouped similar questions together, you'll need to think about the order in which to present those question groups. Most survey researchers agree that it is best to begin a survey with questions that will want to make respondents continue (Babbie, 2010; Dillman, 2000; Neuman, 2003). [3] In other words, don't bore respondents, but don't scare them away either. There's some disagreement over where on a survey to place demographic questions such as those about a person's age, gender, and race. On the one hand, placing them at the beginning of the questionnaire may lead respondents to think the survey is boring, unimportant, and not something they want to bother completing. On the other hand, if your survey deals with some very sensitive or difficult topic, such as child sexual abuse or other criminal activity, you don't want to scare respondents away or shock them by beginning with your most intrusive questions.

In truth, the order in which you present questions on a survey is best determined by the unique characteristics of your research—only you, the researcher, hopefully in consultation with people who are willing to provide you with feedback, can determine how best to order your questions. To do so, think about the unique characteristics of your topic, your questions, and most importantly, your sample. Keeping in mind the characteristics and needs of the people you will ask to complete your survey should help guide you as you determine the most appropriate order in which to present your questions.

You'll also need to consider the time it will take respondents to complete your questionnaire. Surveys vary in length, from just a page or two to a dozen or more pages, which means they also vary in the time it takes to complete them. How long to make your survey depends on several factors. First, what is it that you wish to know? Wanting to understand how grades vary by gender and year in school certainly requires fewer questions than wanting to know how people's experiences in university are shaped by demographic characteristics, university attended, housing situation, fam-

ily background, university major, friendship networks, and extracurricular activities. Keep in mind that even if your research question requires a good number of questions be included in your questionnaire, do your best to keep the questionnaire as brief as possible. Any hint that you've thrown in a bunch of useless questions just for the sake of throwing them in will turn off respondents and may make them not want to complete your survey.

Second, and perhaps more important, how long are respondents likely to be willing to spend completing your questionnaire? If you are studying university students, asking them to use their precious fun time away from studying to complete your survey may mean they won't want to spend more than a few minutes on it. But if you have the endorsement of a professor who is willing to allow you to administer your survey in class, students may be willing to give you a little more time (though perhaps the professor will not). The time that survey researchers ask respondents to spend on questionnaires varies greatly. Some advise that surveys should not take longer than about 15 minutes to complete (Babbie, 2010), others suggest that up to 20 minutes is acceptable (Hopper, 2010). As with question order, there is no clear-cut, always-correct answer about questionnaire length. The unique characteristics of your study and your sample should be considered in order to determine how long to make your questionnaire.

A good way to estimate the time it will take respondents to complete your questionnaire is through pretesting. Pretesting allows you to get feedback on your questionnaire so you can improve it before you actually administer it. Pretesting can be quite expensive and time consuming if you wish to test your questionnaire on a large sample of people who very much resemble the sample to whom you will eventually administer the finalized version of your questionnaire. But you can learn a lot and make great improvements to your questionnaire simply by pretesting with a small number of people to whom you have easy access (perhaps you have a few friends who owe you a favor). By pretesting your questionnaire you can find out how understandable your questions are, get feedback on question wording and order, find out whether any of your questions are exceptionally boring or offensive, and learn whether there are places where you should have included filter questions, to name just a few of the benefits of pretesting. You can also time pretesters as they take your survey. Ask them to complete the survey as though they were actually members of your sample. This will give you a good idea about what sort of time estimate to provide respondents when it comes time to actually administer your survey, and about whether you have some wiggle room to add additional items or need to cut a few items.

Perhaps this goes without saying, but your questionnaire should also be attractive. A messy presentation style can confuse respondents or, at the very least, annoy them. Be brief, to the point, and as clear as possible. Avoid cramming too much into a single page, make your font size readable (at least 12 point), leave a reasonable amount of space between items, and make sure all instructions are exceptionally clear. Think about books, documents, articles, or web pages that you have read yourself—which were relatively easy to read and easy on the eyes and why? Try to mimic those features in the presentation of your survey questions.

Response Rate

It can be very exciting to receive those first few completed surveys back from respondents. Hopefully you'll even get more than a few back, and once you have a handful of completed questionnaires, your feelings may go from initial euphoria to dread. Data are fun and can also be overwhelming. The goal with data analysis is to be able to condense large amounts of information into usable and understandable chunks. Here we'll describe just how that process works for survey researchers.

As mentioned, the hope is that you will receive a good portion of the questionnaires you distributed back in a completed and readable format. The number of completed questionnaires you receive divided by the number of questionnaires you distributed is your ***response rate***. Let's say your sample included 100 people and you sent questionnaires to each of those people. It would be wonderful if all 100 returned completed questionnaires, but the chances of that happening are about zero. If you're lucky, perhaps 75 or so will return completed questionnaires. In this case, your response rate would be 75% (75 divided by 100). That's pretty darn good.

Though response rates vary, and researchers don't always agree about what makes a good response rate, having three-quarters of your surveys returned would be considered good, even excellent, by most survey researchers. There has been lots of research done on how to improve a survey's response rate. Suggestions include personalizing questionnaires by, for example, addressing them to specific respondents rather than to some generic recipient such as "madam" or "sir"; enhancing the questionnaire's credibility by providing details about the study, contact information for the researcher, and perhaps partnering with agencies likely to be respected by respondents such as universities, hospitals, or other relevant organizations; sending out prequestionnaire notices and postquestionnaire reminders; and including some token of appreciation with mailed questionnaires even if small, such as a $1 coin.

The major concern with response rates is that a low rate of response may introduce ***nonresponse bias*** into a study's findings. What if only those who have strong opinions about your study topic return their questionnaires? If that is the case, we may well find that our findings don't at all represent how things really are or, at the very least, we are limited in the claims we can make about patterns found in our data. While high return rates are certainly ideal, a recent body of research shows that concern over response rates may be overblown (Langer, 2003). Several studies have shown that low response rates did not make much difference in findings or in sample representativeness (Curtin, Presser, & Singer, 2000; Keeter, Kennedy, Dimock, Best, & Craighill, 2006; Merkle & Edelman, 2002). For now, the jury may still be out on what makes an ideal response rate and on whether, or to what extent, researchers should be concerned about response rates. Nevertheless, certainly no harm can come from aiming for as high a response rate as possible.

Key Takeaways, Exercises, and References

Key Takeaways

- Strengths of survey research include its cost effectiveness, generalizability, reliability, and versatility.
- Weaknesses of survey research include inflexibility and issues with validity.
- Brainstorming and consulting the literature are two important early steps to take when preparing to write effective survey questions.
- Time is a factor in determining what type of survey researcher administers; cross-sectional surveys are administered at one time, and longitudinal surveys are administered over time.
- Retrospective surveys offer some of the benefits of longitudinal research but also come with their own drawbacks.
- Self-administered questionnaires may be delivered in hard copy form to participants in person or via snail mail or online.
- Make sure that your survey questions will be relevant to all respondents and that you use filter questions when necessary.
- Getting feedback on your survey questions is a crucial step in the process of designing a survey.
- When it comes to creating response options, the solution to the problem of fence-sitting might cause floating, whereas the solution to the problem of floating might cause fence sitting.
- Pretesting is an important step for improving one's survey before actually administering it.
- While survey researchers should always aim to obtain the highest response rate possible, some recent research argues that high return rates on surveys may be less important than we once thought.

Exercises

- What are some ways that survey researchers might overcome the weaknesses of this method?
- Find an article reporting results from survey research (remember how to use Sociological Abstracts?). How do the authors describe the strengths and weaknesses of their study? Are any of the strengths or weaknesses described here mentioned in the article?

- Recall some of the possible research questions you came up with while reading previous chapters of this text. How might you frame those questions so that they could be answered using survey research?
- Do a little Internet research to find out what a Likert scale is and when you may use one.
- Write a closed-ended question that follows the guidelines for good survey question construction. Have a peer in the class check your work (you can do the same for him or her!).
- If the idea of a panel study piqued your interest, check out the Up series of documentary films. While not a survey, the films offer one example of a panel study. Filmmakers began filming the lives of 14 British children in 1964, when the children were 7 years old. They have since caught up with the children every 7 years. In 2012, the eighth installment of the documentary, 56 Up, will come out. Many clips from the series are available on YouTube.
- For more information about online delivery of surveys, check out SurveyMonkey's website: https://www.surveymonkey.com.

References

Babbie, E. (2010). The practice of social research (12th ed.). Belmont, CA: Wadsworth.

Bateman, P. J., Pike, J. C., & Butler, B. S. (2011). To disclose or not: Publicness in social networking sites. Information Technology & People, 24, 78–100.

Belkin, L. (2003, October 26). The opt-out revolution. New York Times, pp. 42–47, 58, 85–86.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. Public Opinion Quarterly, 64, 413–428

Dillman, D. A. (2000). Mail and Internet surveys: The tailored design method (2nd ed.). New York, NY: Wiley;

Hopper, J. (2010). How long should a survey be? Retrieved from https://www.verstaresearch.com/blog/how-long- should-a-survey-be

https://www.aapor.org/Content/aapor/Resources/PollampSurveyFAQ1/DoResponseRatesMatter/Response_Rates_-_Langer.pdf

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. Public Opinion Quarterly, 70, 759–779

Kezdy, A., Martos, T., Boland, V., & Horvath-Szabo, K. (2011). Religious doubts and mental health in adolescence and young adulthood: The association with religious attitudes. Journal of Adolescence, 34, 39–47.

Langer, G. (2003). About response rates: Some unresolved questions. Public Perspective, May/June, 16–18. Retrieved from

Merkle, D. M., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), Survey nonresponse (pp. 243–258). New York, NY: John Wiley and Sons.

Mortimer, J. T. (2003). Working and growing up in America. Cambridge, MA: Harvard University Press.

Neuman, W. L. (2003). Social research methods: Qualitative and quantitative approaches (5th ed.). Boston, MA: Pearson.

Percheski, C. (2008). Opting out? Cohort differences in professional women's employment rates from 1960 to 2005. American Sociological Review, 73, 497–517.

Smith, T. W. (2009). Trends in willingness to vote for a black and woman for president, 1972–2008. GSS Social Change Report No. 55. Chicago, IL: National Opinion Research Center.

CHAPTER 8: EXPERIMENTAL DESIGN

When you think of the term experiment, what comes to mind? Perhaps you thought about trying a new soda or changing your cat's litter to a different brand. We all design informal experiments in our life. We try new things and seek to learn how those things changed us or how they compare to other things we might try. We even create entertainment programs like Mythbusters whose hosts use experimental methods to test whether common myths or bits of folk knowledge are actually true. It's likely you've already developed an intuitive sense of how experiments work. The content of this chapter will increase your existing competency about using experiments to learn about the social world.

LEARNING OBJECTIVES

- Understand when experiments may be appropriate
- Understand the basics of true experimental designs
- Be able to identify "pre-experimental" and "quasi-experimental" designs
- Understand the limitations of various experimental designs
- Be able to explain the logic of experimental design
- Understand threats to the internal and external validity of experimental designs

Chapter 8: Experimental Design is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

Experimental Design: What Is It and When Should It Be Used?

Experiments are an excellent data collection strategy for social workers wishing to observe the effects of a clinical intervention or social welfare program. Understanding what experiments are and how they are conducted is useful for all social scientists, whether they actually plan to use this methodology or simply aim to understand findings from experimental studies. An experiment is a method of data collection designed to test hypotheses under controlled conditions. Students in my research methods classes often use the term experiment to describe all kinds of research projects, but in social scientific research, the term has a unique meaning and should not be used to describe all research methodologies.

Experiments have a long and important history in social science. Behaviorists such as John Watson, B. F. Skinner, Ivan Pavlov, and Albert Bandura used experimental design to demonstrate the various types of conditioning. Using strictly controlled environments, behaviorists were able to isolate a single stimulus as the cause of measurable differences in behavior or physiological responses. The foundations of social learning theory and behavior modification are found in experimental research projects. Moreover, behaviorist experiments brought psychology and social science away from the abstract world of Freudian analysis and towards empirical inquiry, grounded in real-world observations and objectively-defined variables. Experiments are used at all levels of social work inquiry, including agency-based experiments that test therapeutic interventions and policy experiments that test new programs.

Several kinds of experimental designs exist. In general, designs considered to be true experiments contain three key features: independent and dependent variables, pretesting and posttesting, and experimental and control groups. In a true experiment, the effect of an intervention is tested by comparing two groups: one that is exposed to the intervention (the experimental group, also known as the treatment group) and another that does not receive the intervention (the control group).

In some cases, it may be immoral to withhold treatment from a control group within an experiment. If you recruited two groups of people with severe addiction and only provided treatment to one group, the other group would likely suffer. For these cases, researchers use a comparison group that receives "treatment as usual." Experimenters must clearly define what treatment as usual means. For example, a standard treatment in substance abuse recovery is attending Alcoholics Anonymous or Narcotics Anonymous meetings. A substance abuse researcher conducting an experiment may use twelve-step programs in their comparison group and use their experimental intervention in the experimental group. The results would show whether the experimental

intervention worked better than normal treatment, which is useful information. However, using a comparison group is a deviation from true experimental design and is more associated with quasi-experimental designs.

Importantly, participants in a true experiment need to be randomly assigned to either the control or experimental groups. Random assignment uses a random number generator or some other random process to assign people into experimental and control groups. Random assignment is important in experimental research because it helps to ensure that the experimental group and control group are comparable and that any differences between the experimental and control groups are due to random chance. We will address more of the logic behind random assignment in the next section.

In an experiment, the independent variable is the intervention being tested—for example, a therapeutic technique, prevention program, or access to some service or support. It is less common in of social work research, but social science research may also have a stimulus, rather than an intervention as the independent variable. For example, an electric shock or a reading about death might be used as a stimulus to provoke a response.

The dependent variable is usually the intended effect the researcher wants the intervention to have. If the researcher is testing a new therapy for individuals with binge eating disorder, their dependent variable may be the number of binge eating episodes a participant reports. The researcher likely expects her intervention to decrease the number of binge eating episodes reported by participants. Thus, she must measure the number of episodes that existed prior to the intervention, which is the pretest, and after the intervention, which is the posttest.

Let's put these concepts in chronological order so we can better understand how an experiment runs from start to finish. Once you've collected your sample, you'll need to randomly assign your participants to the experimental group and control group. You will then give both groups your pretest, which measures your dependent variable, to see what your participants are like before you start your intervention. Next, you will provide your intervention, or independent variable, to your experimental group. Many interventions last a few weeks or months to complete, particularly therapeutic treatments. Finally, you will administer your posttest to both groups to observer any changes in your dependent variable. Together, this is known as the classic experimental design and is the simplest type of true experimental design. All of the designs we review in this section are variations on this approach.

An interesting example of experimental research can be found in Shannon K. McCoy and Brenda Major's (2003) study of peoples' perceptions of prejudice. In one portion of this multifaceted study, all participants were given a pretest to assess their levels of depression. No significant differences in depression were found between the experimental and control groups during the pretest. Participants in the experimental group were then asked to read an article suggesting that prejudice against their own racial group is severe and pervasive, while participants in the control group were asked to read an article suggesting that prejudice against a racial group other than their own is severe and pervasive. Clearly, these were not meant to be interventions or treatments to help

depression, but were stimuli designed to elicit changes in people's depression levels. Upon measuring depression scores during the posttest period, the researchers discovered that those who had received the experimental stimulus (the article citing prejudice against their same racial group) reported greater depression than those in the control group. This is just one of many examples of social scientific experimental research. In addition to classic experimental design, there are two other ways of designing experiments that are considered to fall within the purview of "true" experiments (Babbie, 2010; Campbell & Stanley, 1963).

The posttest-only control group design is almost the same as classic experimental design, except it does not use a pretest. Researchers who use posttest-only designs want to eliminate testing effects, in which a participant's scores on a measure change because they have already been exposed to it. If you took multiple SAT or ACT practice exams before you took the real one you sent to universities, you've taken advantage of testing effects to get a better score. Considering the previous example on racism and depression, participants who are given a pretest about depression before being exposed to the stimulus would likely assume that the intervention is designed to address depression. That knowledge can cause them to answer differently on the posttest than they otherwise would. Participants are not stupid. They are actively trying to figure out what your study is about.

In theory, as long as the control and experimental groups have been determined randomly and are therefore comparable, no pretest is needed. However, most researchers prefer to use pretests so they may assess change over time within both the experimental and control groups. Researchers wishing to account for testing effects but also gather pretest data can use a Solomon four-group design. In the Solomon four-group design, the researcher uses four groups. Two groups are treated as they would be in a classic experiment—pretest, experimental group intervention, and posttest. The other two groups do not receive the pretest, though one receives the intervention. All groups are given the posttest. Table 12.1 illustrates the features of each of the four groups in the Solomon four-group design. By having one set of experimental and control groups that complete the pretest (Groups 1 and 2) and another set that does not complete the pretest (Groups 3 and 4), researchers using the Solomon four-group design can account for testing effects in their analysis.

Table 8.1: Solomon four-group design

		Pretest	Stimulus	Posttest
random assignment	**Group 1**	X	X	X
	Group 2	X		X
	Group 3		X	X
	Group 4			X

Solomon four-group designs are challenging to implement in the real world because they are time- and resource-intensive. Researchers must recruit enough participants to create four groups and implement interventions in two of them. Overall, true experimental designs are sometimes difficult to implement in a real-world practice environment. It may be impossible to withhold treatment from a control group or randomly assign participants in a study. In these cases, pre- experimental and quasi-experimental designs can be used. However, the differences in rigor from true experimental designs leave their conclusions more open to critique.

Chapter 8: Experimental Design is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

Pre-experimental and Quasi-Experimental Design

As we discussed in the previous section, time, funding, and ethics may limit a researcher's ability to conduct a true experiment. For researchers in the medical sciences and social work, conducting a true experiment could require denying needed treatment to clients, which is a clear ethical violation. Even those whose research may not involve the administration of needed medications or treatments may be limited in their ability to conduct a classic experiment. When true experiments are not possible, researchers often use quasi-experimental designs.

Quasi-experimental designs are similar to true experiments, but they lack random assignment to experimental and control groups. The most basic of these quasi-experimental designs is the nonequivalent comparison groups design (Rubin & Babbie, 2017). The nonequivalent comparison group design looks a lot like the classic experimental design, except it does not use random assignment. In many cases, these groups may already exist. For example, a researcher might conduct research at two different agency sites, one of which receives the intervention and the other does not. No one was assigned to treatment or comparison groups. Those groupings existed prior to the study. While this method is more convenient for real-world research, researchers cannot be sure that the groups are comparable. Perhaps the treatment group has a characteristic that is unique–for example, higher income or different diagnoses–that make the treatment more effective.

Quasi-experiments are particularly useful in social welfare policy research. Social welfare policy researchers like me often look for what are termed natural experiments, or situations in which comparable groups are created by differences that already occur in the real world. For example, Stratmann and Wille (2016) were interested in the effects of a state healthcare policy called Certificate of Need on the quality of hospitals. They clearly cannot assign states to adopt one set of policies or another. Instead, researchers used hospital referral regions, or the areas from which hospitals draw their patients, that spanned across state lines. Because the hospitals were in the same referral region, researchers could be pretty sure that the client characteristics were pretty similar. In this way, they could classify patients in experimental and comparison groups without affecting policy or telling people where to live.

There are important examples of policy experiments that use random assignment, including the Oregon Medicaid experiment. In the Oregon Medicaid experiment, the wait list for Oregon was so long, state officials conducted a lottery to see who from the wait list would receive Medicaid (Baicker et al., 2013). Researchers used the lottery as a natural experiment that included random assignment. People selected to be a part of Medicaid were the experimental group and those on the wait list were in the control group. There are some practical complications with using people on a wait list as a control group—most obviously, what happens when people on the wait list are accepted into the program while you're still collecting data? Natural experiments aren't a specific

kind of experiment like quasi- or pre-experimental designs. Instead, they are more like a feature of the social world that allows researchers to use the logic of experimental design to investigate the connection between variables.

Matching is another approach in quasi-experimental (and true experimental) design to assigning experimental and comparison groups. Researchers should think about what variables are important in their study, particularly demographic variables or attributes that might impact their dependent variable. Individual matching involves pairing participants with similar attributes. When this is done at the beginning of an experiment, the matched pair is split—with one participant going to the experimental group and the other to the control group. An ex post facto control group, in contrast, is when a researcher matches individuals after the intervention is administered to some participants. Finally, researchers may engage in aggregate matching, in which the comparison group is determined to be similar on important variables.

There are many different quasi-experimental designs in addition to the nonequivalent comparison group design described earlier. Describing all of them is beyond the scope of this textbook, but one more design is worth mentioning. The time series design uses multiple observations before and after an intervention. In some cases, experimental and comparison groups are used. In other cases where that is not feasible, a single experimental group is used. By using multiple observations before and after the intervention, the researcher can better understand the true value of the dependent variable in each participant before the intervention starts. Additionally, multiple observations afterwards allow the researcher to see whether the intervention had lasting effects on participants. Time series designs are similar to single-subjects designs, which we will discuss in Chapter 15.

When true experiments and quasi-experiments are not possible, researchers may turn to a pre-experimental design (Campbell & Stanley, 1963). Pre-experimental designs are called such because they often happen before a true experiment is conducted. Researchers want to see if their interventions will have some effect on a small group of people before they seek funding and dedicate time to conduct a true experiment. Pre-experimental designs, thus, are usually conducted as a first step towards establishing the evidence for or against an intervention. However, this type of design comes with some unique disadvantages, which we'll describe as we review the pre-experimental designs available.

If we wished to measure the impact of a natural disaster, such as Hurricane Katrina for example, we might conduct a pre-experiment by identifying an experimental group from a community that experienced the hurricane and a control group from a similar community that had not been hit by the hurricane. This study design, called a static group comparison, has the advantage of including a comparison group that did not experience the stimulus (in this case, the hurricane). Unfortunately, it is difficult to know those groups are truly comparable because the experimental and control groups were determined by factors other than random assignment. Additionally, the design would only allow for posttests, unless one were lucky enough to be gathering the data already before Kat-

rina. As you might have guessed from our example, static group comparisons are useful in cases where a researcher cannot control or predict whether, when, or how the stimulus is administered, as in the case of natural disasters.

In cases where the administration of the stimulus is quite costly or otherwise not possible, a one-shot case study design might be used. In this instance, no pretest is administered, nor is a control group present. In our example of the study of the impact of Hurricane Katrina, a researcher using this design would test the impact of Katrina only among a community that was hit by the hurricane and would not seek a comparison group from a community that did not experience the hurricane. Researchers using this design must be extremely cautious about making claims regarding the effect of the stimulus, though the design could be useful for exploratory studies aimed at testing one's measures or the feasibility of further study.

Finally, if a researcher is unlikely to be able to identify a sample large enough to split into control and experimental groups, or if she simply doesn't have access to a control group, the researcher might use a one-group pre-/posttest design. In this instance, pre- and posttests are both taken, but there is no control group to which to compare the experimental group. We might be able to study of the impact of Hurricane Katrina using this design if we'd been collecting data on the impacted communities prior to the hurricane. We could then collect similar data after the hurricane. Applying this design involves a bit of serendipity and chance. Without having collected data from impacted communities prior to the hurricane, we would be unable to employ a one- group pre-/posttest design to study Hurricane Katrina's impact.

As implied by the preceding examples where we considered studying the impact of Hurricane Katrina, experiments do not necessarily need to take place in the controlled setting of a lab. In fact, many applied researchers rely on experiments to assess the impact and effectiveness of various programs and policies. You might recall our discussion of arresting perpetrators of domestic violence in Chapter 6, which is an excellent example of an applied experiment. Researchers did not subject participants to conditions in a lab setting; instead, they applied their stimulus (in this case, arrest) to some subjects in the field and they also had a control group in the field that did not receive the stimulus (and therefore were not arrested).

Chapter 8: Experimental Design is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

The Logic of Experimental Design

As we discussed at the beginning of this chapter, experimental design is commonly understood and implemented informally in everyday life. Trying out a new restaurant, dating a new person—we often term these experiments. As you've learned over the past two sections, in order for something to be a true experiment, or even a quasi- or pre-experiment, you must rigorously apply the various components of experimental design. A true experiment for trying a new restaurant would include recruitment of a large enough sample, random assignment to control and experimental groups, pretesting and posttesting, as well as using clearly and objectively defined measures of satisfaction with the restaurant.

Social scientists use this level of rigor and control because they try to maximize the internal validity of their experiment. Internal validity is the confidence researchers have about whether their intervention produced variation in their dependent variable. Thus, experiments are attempts to establish causality between two variables—your treatment and its intended outcome. Causal relationships must establish four criteria: covariation, plausibility, temporality, and non-spuriousness.

The logic and rigor experimental design allows for causal relationships to be established. Experimenters can assess covariation on the dependent variable through pre- and posttests. The use of experimental and control conditions ensures that some people receive the intervention and others do not, providing variation in the independent variable. Moreover, since the researcher controls when the intervention is administered, she can be assured that changes in the independent variable (the treatment) happened before changes the dependent variable (the outcome). In this way, experiments assure temporality. In our restaurant experiment, we would know through assignment experimental and control groups that people varied in the restaurant they attended. We would also know whether their level of satisfaction changed, as measured by the pre- and posttest. We would also know that changes in our diners' satisfaction occurred after they left the restaurant, not before they walked in because of the pre- and posttest.

Experimenters will also have a plausible reason why their intervention would cause changes in the dependent variable. Usually, a theory or previous empirical evidence should indicate the potential for a causal relationship. Perhaps we found a national poll that found the type of food our experimental restaurant served, let's say pizza, is the most popular food in America. Perhaps this restaurant has good reviews on Yelp or Google. This evidence would give us a plausible reason to establish our restaurant as causing satisfaction.

One of the most important features of experiments is that they allow researchers to eliminate spurious variables. True experiments are usually conducted under strictly controlled laboratory conditions. The intervention must be given in the same way to each person, with a minimal number of other variables that might cause their posttest scores to change. In our restaurant example, this level of control might prove difficult. We cannot control how many people are waiting for a table,

whether participants saw someone famous there, or if there is bad weather. Any of these factors might cause a diner to be less satisfied with their meal. These spurious variables may cause changes in satisfaction that have nothing to do with the restaurant itself, an important problem in real-world research. For this reason, experiments use the laboratory environment try to control as many aspects of the research process as possible. Researchers in large experiments often employ clinicians or other research staff to help them. Researchers train their staff members exhaustively, provide pre-scripted responses to common questions, and control the physical environment of the lab so each person who participates receives the exact same treatment.

Experimental researchers also document their procedures, so that others can review how well they controlled for spurious variables. A good example of this concept is Bruce Alexander's Rat Park (1981) experiments. Much of the early research conducted on addictive drugs, like heroin and cocaine, was conducted on animals other than humans, usually mice or rats. While this may seem strange, the systems of our mammalian relatives are similar enough to humans that causal inferences can be made from animal studies to human studies. It is certainly unethical to deliberately cause humans to become addicted to cocaine and measure them for weeks in a laboratory, but it is currently more ethically acceptable to do so with animals. There are specific ethical processes for animal research, similar to an IRB review.

The scientific consensus up until Alexander's experiments was that cocaine and heroin were so addictive that rats, if offered the drugs, would consume them repeatedly until they perished. Researchers claimed this behavior explained how addiction worked in humans, but Alexander was not so sure. He knew rats were social animals and the experimental procedure from previous experiments did not allow them to socialize. Instead, rats were kept isolated in small cages with only food, water, and metal walls. To Alexander, social isolation was a spurious variable, causing changes in addictive behavior not due to the drug itself. Alexander created an experiment of his own, in which rats were allowed to run freely in an interesting environment, socialize and mate with other rats, and of course, drink from a solution that contained an addictive drug. In this environment, rats did not become hopelessly addicted to drugs. In fact, they had little interest in the substance.

To Alexander, the results of his experiment demonstrated that social isolation was more of a causal factor for addiction than the drug itself. This makes intuitive sense to me. If I were in solitary confinement cell for most of my life, the escape of an addictive drug would seem more tempting than if I were in my natural environment with friends, family, and activities. One challenge with Alexander's findings is that subsequent researchers have had mixed success replicating his findings (e.g., Petrie, 1996; Solinas, Thiriet, El Rawas, Lardeux, & Jaber, 2009). Replication involves conducting another researcher's experiment in the same manner and seeing if it produces the same results. If the causal relationship is real, it should occur in all (or at least most) replications of the experiment.

One of the defining features of experiments is that they report their procedures diligently, which allows for easier replication. Recently, researchers at the Reproducibility Project have caused a significant controversy in social science fields like psychology (Open Science Collaboration, 2015). In one study, researchers attempted reproduce the results of 100 experiments published in major psychology journals between 2008 and the present. What they found was shocking. The results of only 36% of the studies were reproducible. Despite coordinating closely with the original researchers, the Reproducibility Project found that nearly two-thirds of psychology experiments published in respected journals were not reproducible. The implications of the Reproducibility Project are staggering, and social scientists are coming up with new ways to ensure researchers do not cherry-pick data or change their hypotheses, simply to get published.

Returning to Alexander's Rat Park study, consider what the implications of his experiment were to a substance abuse professional. The conclusions he drew from his experiments on rats were meant to generalize to the population of people with substance use disorders. Experiments seek to establish external validity, which is the degree to which their conclusions generalize to larger populations and different situations. Alexander argues his conclusions about addiction and social isolation help us understand why people living in deprived, isolated environments will often become addicted to drugs more often than those in more enriching environments. Similarly, earlier rat researchers argued their results showed these drugs were instantly addictive, often to the point of death.

Neither study will match up perfectly with real life. One may encounter many individuals who may have fit into Alexander's social isolation model, but social isolations for humans is complex. If individuals live in environments with other sociable humans, work jobs, and have romantic relationships, how isolated are they? On the other hand, many may face structural racism, poverty, trauma, and other challenges that may contribute to social isolation. Alexander's work helps us understand some of these experiences, but the explanation was incomplete. The real world is much more complicated than the experimental conditions in Rat Park, just as humans are more complex than rats.

Social scientists are especially attentive to how social context shapes social life. So, we are likely to point out a specific disadvantage of experiments. They are rather artificial. How often do real-world social interactions occur in the same way that they do in a lab? Experiments that are conducted in community settings may not be as subject to artificiality, though then their conditions are less easily controlled. This relationship demonstrates the tension between internal and external validity. The more researchers tightly control the environment to ensure internal validity, the less they can claim external validity and that their results are applicable to different populations and circumstances. Correspondingly, researchers whose settings are just like the real world will be less able to ensure internal validity, as there are many factors that could pollute the research process. This is not to suggest that experimental research cannot have external validity, but experimental researchers must always be aware that external validity problems can occur and be forthcoming in their reports of findings about this potential weakness.

Chapter 8: Experimental Design is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

Threats to Validity

Internal validity and external validity are conceptually linked. Internal validity refers to the degree to which the intervention causes its intended outcomes, and external validity refers to how well that relationship applies to different groups and circumstances. There are a number of factors that may influence a study's validity. You might consider these threats to all be spurious variables, as we discussed at the beginning of this section. Each threat proposes another factor that is changing the relationship between intervention and outcome. The threats introduce error and bias into the experiment.

Throughout this chapter, we reviewed the importance of experimental and control groups. These groups must be comparable in order for experimental design to work. Comparable groups are groups that are similar across factors important for the study. Researchers can help establish comparable groups by using probability sampling, random assignment, or matching techniques. Control or comparison groups provide a counterfactual—what would have happened to my experimental group had I not given them my intervention? Two very different groups would not allow you to answer that question. Intuitively, we all know that no two people are exactly the same. So, no groups are ever perfectly comparable. What's important is ensuring groups are comparable along the variables relevant to the research project.

In our restaurant example, if one of my groups had far more vegetarians or people with gluten issues, it might influence how satisfied they were with my restaurant. My groups, in that case, would not be comparable. Researchers also account for this by measuring other variables, like dietary preference, and controlling for their effects statistically, after the data are collected. Similarly, if I were to pick out people I thought would "really like" my restaurant and assign them to the experimental group, I would be introducing selection bias into my sample. This is another reason experimenters use random assignment, so conscious and unconscious bias do not influence to which group a participant is assigned.

Experimenters themselves are often the source of threats to validity. They may choose measures that do not accurately measure participants or implement the measure in a way that biases participant responses in one direction or another. Researchers may, just by the very act of conducting an experiment, influence participants to perform differently. Experiments are different from participants' normal routines. The novelty of a research environment or experimental treatment may cause them to expect to feel differently, independently of the actual intervention. You have likely heard of the placebo effect, in which a participant feels better, despite having received no intervention at all.

Researchers may also introduce error by expecting participants in each group to behave differently. For the experimental group, researchers may expect them to feel better and may give off conscious or unconscious cues to participants that influence their outcomes. Control groups will be expected to fare worse, and research staff could cue participants that they should feel worse

than they otherwise would. For this reason, researchers often use double-blind designs wherein research staff interacting with participants are unaware of who is in the control or experimental group. Proper training and supervision are also necessary to account for these and other threats to validity. If proper supervision is not applied, research staff administering the control group may try to equalize treatment or engage in a rivalry with research staff administering the experimental group (Engel & Schutt, 2016).

No matter how tightly the researcher controls the experiment, participants are humans and are therefore curious, problem-solving creatures. Participants who learn they are in the control group may react by trying to outperform the experimental group or by becoming demoralized. In either case, their outcomes in the study would be different had they been unaware of their group assignment. Participants in the experimental group may begin to behave differently or share insights from the intervention with individuals in the control group. Whether through social learning or conversation, participants in the control group may receive parts of the intervention of which they were supposed to be unaware. Experimenters, as a result, try to keep experimental and control groups as separate as possible. Inside a laboratory study, this is significantly easier as the researchers control access and timing at the facility. In agency-based research, this problem is more complicated. If your intervention is good, your participants in the experimental group may impact the control group by behaving differently and sharing the insights they've learned with their peers. Agency-based researchers may locate experimental and control conditions at separate offices with separate treatment staff to minimize the interaction between their participants.

Chapter 8: Experimental Design is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

Key Takeaways, Exercises, and References

Key Takeaways

- An experiment is a type of empirical study that features the manipulation of an independent variable, the measurement of a dependent variable, and control of extraneous variables.
- An extraneous variable is any variable other than the independent and dependent variables. A confound is an extraneous variable that varies systematically with the independent variable.
- Experimental research on the effectiveness of a treatment requires both a treatment condition and a control condition, which can be a no-treatment control condition, a placebo control condition, or a wait- list control condition. Experimental treatments can also be compared with the best available alternative.
- Experiments can be conducted using either between-subjects or within-subjects designs. Deciding which to use in a particular situation requires careful consideration of the pros and cons of each approach.
- Random assignment to conditions in between-subjects experiments or counterbalancing of orders of conditions in within-subjects experiments is a fundamental element of experimental research. The purpose of these techniques is to control extraneous variables so that they do not become confounding variables.
- Studies are high in internal validity to the extent that the way they are conducted supports the conclusion that the independent variable caused any observed differences in the dependent variable. Experiments are generally high in internal validity because of the manipulation of the independent variable and control of extraneous variables.
- Studies are high in external validity to the extent that the result can be generalized to people and situations beyond those actually studied. Although experiments can seem "artificial"—and low in external validity—it is important to consider whether the psychological processes under study are likely to operate in other people and situations.
- There are several effective methods you can use to recruit research participants for your experiment, including through formal subject pools, advertisements, and personal appeals. Field experiments require well-defined participant selection procedures.
- It is important to standardize experimental procedures to minimize extraneous variables, including experimenter expectancy effects.
- It is important to conduct one or more small-scale pilot tests of an experiment to be sure that the procedure works as planned.

Exercises

For each of the following, identify and design an appropriate experiment. What threats to validity are there and how have you sought to minimize these?

- You want to test the relative effectiveness of two training programs for running a marathon.
- Using photographs of people as stimuli, you want to see if smiling people are perceived as more intelligent than people who are not smiling.
- You want to see if the way a panhandler is dressed (neatly vs. sloppily) affects whether or not passersby give him any money.

References

A. N. (2013). The Oregon experiment—effects of Medicaid on clinical outcomes. New England Journal of Medicine, 368(18), 1713-1722.

Alexander, B. (2010). Addiction: The view from rat park. Retrieved from: https://www.brucekalexander.com/articles-speeches/rat-park/148-addiction-the-view-from-rat- park

Babbie, E. (2010). The practice of social research (12th ed.). Belmont, CA: Wadsworth;

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., ... & Finkelstein,

Campbell, D., & Stanley, J. (1963). Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally.

Campbell, D., & Stanley, J. (1963). Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally.

Engel, R. J. & Schutt, R. K. (2016). The practice of research in social work (4th ed.). Washington, DC: SAGE Publishing.

McCoy, S. K., & Major, B. (2003). Group identification moderates emotional response to perceived prejudice. Personality and Social Psychology Bulletin, 29, 1005–1017.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.

Petrie, B. F. (1996). Environment is not the most important variable in determining oral morphine consumption in Wistar rats. Psychological reports, 78(2), 391-400.

Rubin, C. & Babbie, S. (2017). Research methods for social work (9th edition). Boston, MA: Cengage.

Solinas, M., Thiriet, N., El Rawas, R., Lardeux, V., & Jaber, M. (2009). Environmental enrichment during early stages of life reduces the behavioral, neurochemical, and molecular effects of cocaine. Neuropsychopharmacology, 34(5), 1102.

Stratmann, T. & Wille, D. (2016). Certificate-of-need laws and hospital quality. Mercatus Center at George Mason University, Arlington, VA. Retrieved from: https://www.mercatus.org/system/files/ mercatus-stratmann-wille-con-hospital-quality-v1.pdf

Chapter 8: Experimental Design is adapted from Matthew DeCarlo (2018) ***Scientific Inquiry in Social Work*** and is licensed under a **CC BY-NC-SA** Licence.

CHAPTER 9: DESCRIPTIVE STATISTICS

At this point, we need to consider the basics of data analysis in social scientific research in more detail. In this chapter, we focus on descriptive statistics—a set of techniques for summarizing and displaying the data from your sample. We look first at some of the most common techniques for describing single variables, followed by some of the most common techniques for describing statistical relationships between variables. We then look at how to present descriptive statistics in writing and also in the form of tables and graphs that would be appropriate for an American Psychological Association (APA)-style research report. We end with some practical advice for organizing and carrying out your analyses.

LEARNING OBJECTIVES

- Understand the characteristics of distributions, including shape, modality, and symmetry
- Understand the measures of central tendency and under what conditions each may be used
- Understand the measures of dispersion of a distribution of scores
- Be aware how Cohen's d and Pearson's r demonstrate relationships among variables
- Be aware of common guidelines for the presentation of data in text, including in parentheses, in a table, or in a figure
- Understand why descriptive statistics are essential to quantitative analysis

Chapter 9: Descriptive Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Describing Single Variables

Descriptive statistics refers to a set of techniques for summarizing and displaying data. Let us assume here that the data are quantitative and consist of scores on one or more variables for each of several study participants. Although in most cases the primary research question will be about one or more statistical relationships between variables, it is also important to describe each variable individually. For this reason, we begin by looking at some of the most common techniques for describing single variables.

The Distribution of a Variable

Every variable has a distribution, which is the way the scores are distributed across the levels of that variable. For example, in a sample of 100 university students, the distribution of the variable "number of siblings" might be such that 10 of them have no siblings, 30 have one sibling, 40 have two siblings, and so on. In the same sample, the distribution of the variable "sex" might be such that 44 have a score of "male" and 56 have a score of "female."

Frequency Tables

One way to display the distribution of a variable is in a frequency table. Table 12.1, for example, is a frequency table showing a hypothetical distribution of scores on the Rosenberg Self-Esteem Scale for a sample of 40 university students. The first column lists the values of the variable—the possible scores on the Rosenberg scale—and the second column lists the frequency of each score. This table shows that there were three students who had self-esteem scores of 24, five who had self-esteem scores of 23, and so on. From a frequency table like this, one can quickly see several important aspects of a distribution, including the range of scores (from 15 to 24), the most and least common scores (22 and 17, respectively), and any extreme scores that stand out from the rest.

Table 9.1: Hypothetical Frequency Distribution of Scores on the Rosenberg Self-Esteem Scale

Self-esteem	Frequency
24	3
23	5
22	10
21	8
20	5
19	3
18	3
17	0
16	2
15	1

There are a few other points worth noting about frequency tables. First, the levels listed in the first column usually go from the highest at the top to the lowest at the bottom, and they usually do not extend beyond the highest and lowest scores in the data. For example, although scores on the Rosenberg scale can vary from a high of 30 to a low of 0, Table 12.1 only includes levels from 24 to 15 because that range includes all the scores in this particular data set. Second, when there are many different scores across a wide range of values, it is often better to create a grouped frequency table, in which the first column lists ranges of values and the second column lists the frequency of scores in each range. Table 12.2, for example, is a grouped frequency table showing a hypothetical distribution of simple reaction times for a sample of 20 participants. In a grouped frequency table, the ranges must all be of equal width, and there are usually between five and 15 of them.

Finally, frequency tables can also be used for categorical variables, in which case the levels are category labels. The order of the category labels is somewhat arbitrary, but they are often listed from the most frequent at the top to the least frequent at the bottom.

Table 9.2: Hypothetical Distribution of Reaction Times

Reaction time (ms)	Frequency
241–260	1
221–240	2
201–220	2
181–200	9
161–180	4
141–160	2

Histograms

A histogram is a graphical display of a distribution. It presents the same information as a frequency table but in a way that is even quicker and easier to grasp. The histogram in Figure 12.1 presents the distribution of self- esteem scores in Table 12.1. The x-axis of the histogram represents the variable and the y-axis represents frequency. Above each level of the variable on the x-axis is a vertical bar that represents the number of individuals with that score. When the variable is quantitative, as in this example, there is usually no gap between the bars. When the variable is categorical, however, there is usually a small gap between them. (The gap at 17 in this histogram reflects the fact that there were no scores of 17 in this data set.)

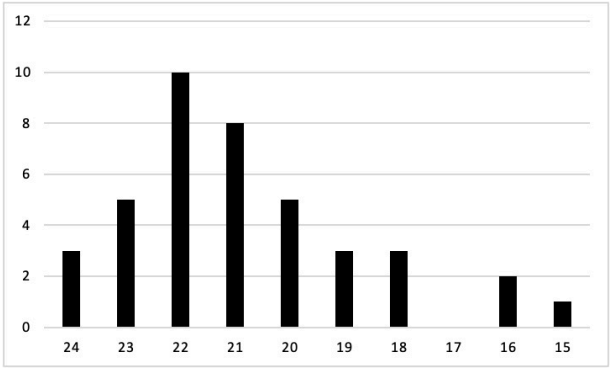


Figure 9.1 Histogram Showing the Distribution of Self- Esteem Scores Presented in Table 9.1

Distribution Shapes

When the distribution of a quantitative variable is displayed in a histogram, it has a shape. The shape of the distribution of self-esteem scores in Figure 12.1 is typical. There is a peak somewhere near the middle of the distribution and "tails" that taper in either direction from the peak. The distribution of Figure 12.1 is unimodal, meaning it has one distinct peak, but distributions can also be bimodal, meaning they have two distinct peaks. Figure 12.2, for example, shows a hypothetical bimodal distribution of scores on the Beck Depression Inventory. Distributions can also have more than two distinct peaks, but these are relatively rare in psychological research.

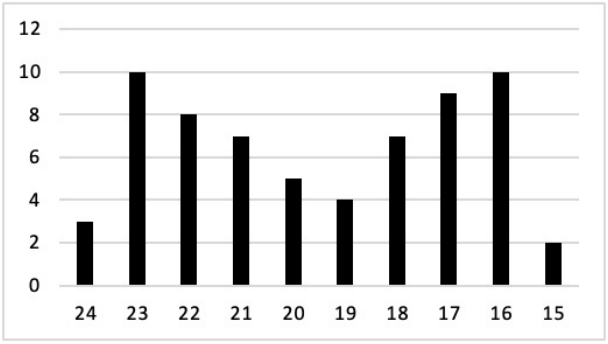


Figure 9.2 Histogram Illustrating a Bi-Modal Distribution

Another characteristic of the shape of a distribution is whether it is symmetrical or skewed. The distribution in the center of Figure 12.3 is symmetrical. Its left and right halves are mirror images of each other. The distribution on the left is negatively skewed, with its peak shifted toward the upper end of its range and a relatively long negative tail. The distribution on the right is positively skewed, with its peak toward the lower end of its range and a relatively long positive tail.

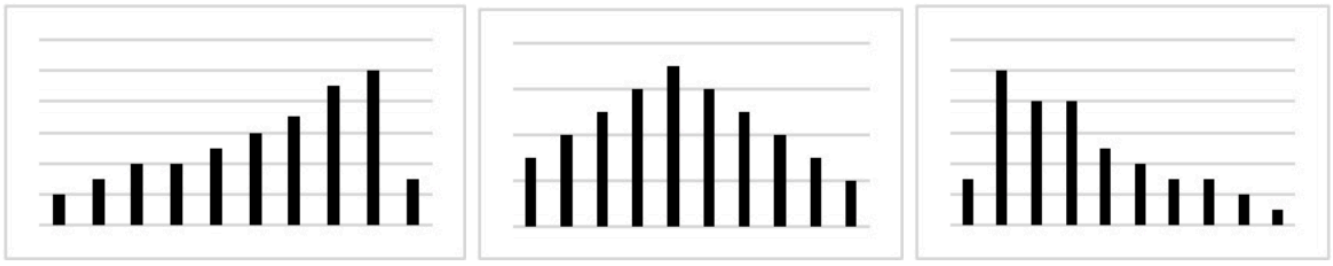


Figure 9.3 Negatively Skewed, Symmetrical, and Positively Skewed Distributions

An outlier is an extreme score that is much higher or lower than the rest of the scores in the distribution. Sometimes outliers represent truly extreme scores on the variable of interest. For example, on the Beck Depression Inventory, a single clinically depressed person might be an outlier in a sample of otherwise happy and high-functioning peers. However, outliers can also represent errors or misunderstandings on the part of the researcher or participant, equipment malfunctions, or similar problems. We will say more about how to interpret outliers and what to do about them later in this chapter.

Chapter 9: Descriptive Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Measures of Central Tendency and Variability

It is also useful to be able to describe the characteristics of a distribution more precisely. Here we look at how to do this in terms of two important characteristics: their central tendency and their variability.

Central Tendency

The central tendency of a distribution is its middle—the point around which the scores in the distribution tend to cluster. (Another term for central tendency is average.) Looking back at Figure 9.1, for example, we can see that the self-esteem scores tend to cluster around the values of 20 to 22. Here we will consider the three most common measures of central tendency: the mean, the median, and the mode.

The **mean** of a distribution (symbolized M) is the sum of the scores divided by the number of scores. It is an average. As a formula, it looks like this:

$$M=\Sigma X/N$$

In this formula, the symbol Σ (the Greek letter sigma) is the summation sign and means to sum across the values of the variable X. N represents the number of scores. The mean is by far the most common measure of central tendency, and there are some good reasons for this. It usually provides a good indication of the central tendency of a distribution, and it is easily understood by most people. In addition, the mean has statistical properties that make it especially useful in doing inferential statistics.

An alternative to the mean is the median. The **median** is the middle score in the sense that half the scores in the distribution are less than it and half are greater than it. The simplest way to find the median is to organize the scores from lowest to highest and locate the score in the middle. Consider, for example, the following set of seven scores:

8 4 12 14 3 2 3

To find the median, simply rearrange the scores from lowest to highest and locate the one in the middle.

2 3 3 4 8 12 14

In this case, the median is 4 because there are three scores lower than 4 and three scores higher than 4.

When there is an even number of scores, there are two scores in the middle of the distribution, in which case the median is the value halfway between them. For example, if we were to add a score of 15 to the preceding data set, there would be two scores (both 4 and 8) in the middle of the distribution, and the median would be halfway between them (6).

One final measure of central tendency is the mode. The **mode** is the most frequent score in a distribution. In the self-esteem distribution presented in Table 12.1 and Figure 12.1, for example, the mode is 22. More students had that score than any other. The mode is the only measure of central tendency that can also be used for categorical variables.

In a distribution that is both unimodal and symmetrical, the mean, median, and mode will be very close to each other at the peak of the distribution. In a bimodal or asymmetrical distribution, the mean, median, and mode can be quite different. In a bimodal distribution, the mean and median will tend to be between the peaks, while the mode will be at the tallest peak. In a skewed distribution, the mean will differ from the median in the direction of the skew (i.e., the direction of the longer tail). For highly skewed distributions, the mean can be pulled so far in the direction of the skew that it is no longer a good measure of the central tendency of that distribution. Imagine, for example, a set of four simple reaction times of 200, 250, 280, and 250 milliseconds (ms). The mean is 245 ms. But the addition of one more score of 5,000 ms—perhaps because the participant was not paying attention—would raise the mean to 1,445 ms. Not only is this measure of central tendency greater than 80% of the scores in the distribution, but it also does not seem to represent the behavior of anyone in the distribution very well. This is why researchers often prefer the median for highly skewed distributions (such as distributions of reaction times).

Keep in mind, though, that you are not required to choose a single measure of central tendency in analyzing your data. Each one provides slightly different information, and all of them can be useful.

Measures of Variability

The variability of a distribution is the extent to which the scores vary around their central tendency. Consider the two distributions in Figure 9.4, each of which has the same central tendency. The mean, median, and mode of each distribution are 5. Notice, however, that the distributions differ in terms of their variability. From left to right in Figure 9.4, the observations become less clustered around the centre value.

One simple measure of variability is the ***range***, which is simply the difference between the highest and lowest scores in the distribution. The range of the self-esteem scores in Table 9.1, for example, is the difference between the highest score (24) and the lowest score (15). That is, the range is 24 − 15 = 9. Although the range is easy to compute and understand, it can be misleading when there are outliers. Imagine, for example, an exam on which all the students scored between 90 and 100.

It has a range of 10. But if there was a single student who scored 20, the range would increase to 80—giving the impression that the scores were quite variable when in fact only one student differed substantially from the rest.

By far the most common measure of variability is the standard deviation. The ***standard deviation*** of a distribution is the average distance between the scores and the mean. The broader the distribution (the more observations diverge from the mean), the higher the standard deviation and the narrower the distribution (the less observations diverge from the mean), the smaller the standard deviation..

Computing the standard deviation involves a slight complication. Specifically, it involves finding the difference between each score and the mean, squaring each difference, finding the mean of these squared differences, and finally finding the square root of that mean. The formula looks like this:

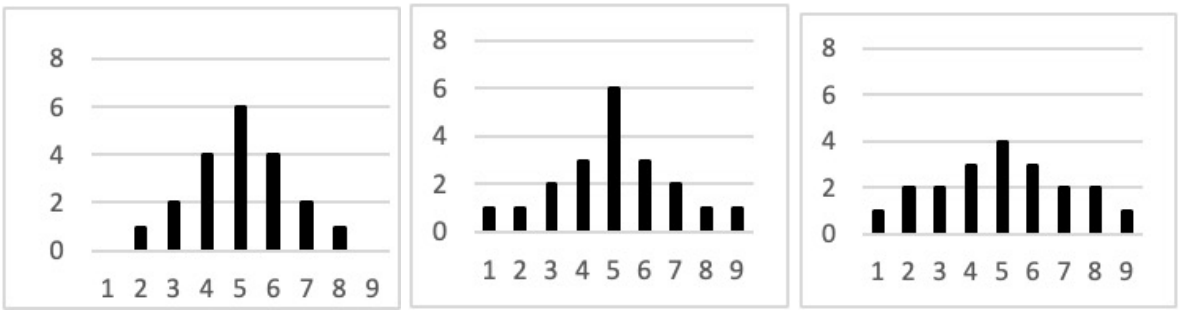


Figure 9.4 Distributions With the Same Mean, Median and Mode and Varying Standard Deviation

Percentile Ranks and z Scores

In many situations, it is useful to have a way to describe the location of an individual score within its distribution. One approach is the percentile rank. The ***percentile rank*** of a score is the percentage of scores in the distribution that are lower than that score. Consider, for example, the distribution in Table 12.1. For any score in the distribution, we can find its percentile rank by counting the number of scores in the distribution that are lower than that score and converting that number to a percentage of the total number of scores.

Notice, for example, that five of the students represented by the data in Table 9.1 had self-esteem scores of 23. In this distribution, 32 of the 40 scores (80%) are lower than 23. Thus each of these students has a percentile rank of 80. (It can also be said that they scored "at the 80th percentile.") Percentile ranks are often used to report the results of standardized tests of ability or achievement. If your percentile rank on a test of verbal ability were 40, for example, this would mean that you scored higher than 40% of the people who took the test.

Another approach is the z score. The ***z score*** for a particular individual is the difference between that individual's score and the mean of the distribution, divided by the standard deviation of the distribution:

$$z = (X-M)/SD$$

A z score indicates how far above or below the mean a raw score is, but it expresses this in terms of the standard deviation. For example, in a distribution of intelligence quotient (IQ) scores with a mean of 100 and a standard deviation of 15, an IQ score of 110 would have a z score of (110 − 100) / 15 = +0.67. In other words, a score of 110 is 0.67 standard deviations (approximately two thirds of a standard deviation) above the mean.

Similarly, a raw score of 85 would have a z score of (85 − 100) / 15 = −1.00. In other words, a score of 85 is one standard deviation below the mean.

There are several reasons that z scores are important. Again, they provide a way of describing where an individual's score is located within a distribution and are sometimes used to report the results of standardized tests. They also provide one way of defining outliers. For example, outliers are sometimes defined as scores that have z scores less than −3.00 or greater than +3.00. In other words, they are defined as scores that are more than three standard deviations from the mean. Finally, z scores play an important role in understanding and computing other statistics, as we will see shortly.

Chapter 9: Descriptive Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Describing Statistical Relationships

As we have seen throughout this book, most interesting research questions in quantitative social sciences are about statistical relationships between variables. In this section, we revisit the two basic forms of statistical relationship introduced earlier in the book—differences between groups or conditions and relationships between quantitative variables—and we consider how to describe them in more detail.

Differences Between Groups or Conditions

Differences between groups or conditions are usually described in terms of the mean and standard deviation of each group or condition. For example, Thomas Ollendick and his colleagues conducted a study in which they evaluated two one-session treatments for simple phobias in children (Ollendick et al., 2009). They randomly assigned children with an intense fear (e.g., to dogs) to one of three conditions. In the exposure condition, the children actually confronted the object of their fear under the guidance of a trained therapist. In the education condition, they learned about phobias and some strategies for coping with them. In the wait-list control condition, they were waiting to receive a treatment after the study was over. The severity of each child's phobia was then rated on a 1-to-8 scale by a clinician who did not know which treatment the child had received. (This was one of several dependent variables.) The mean fear rating in the education condition was 4.83 with a standard deviation of 1.52, while the mean fear rating in the exposure condition was 3.47 with a standard deviation of 1.77. The mean fear rating in the control condition was 5.56 with a standard deviation of 1.21. In other words, both treatments worked, but the exposure treatment worked better than the education treatment. As we have seen, differences between group or condition means can be presented in a bar graph.

It is also important to be able to describe the strength of a statistical relationship, which is often referred to as the ***effect size***. The most widely used measure of effect size for differences between group or condition means is called Cohen's d, which is the difference between the two means divided by the standard deviation:

$$d = (M1 - M2)/SD$$

Conceptually, Cohen's d is the difference between the two means expressed in standard deviation units. (Notice its similarity to a z score, which expresses the difference between an individual score and a mean in standard deviation units.) A Cohen's d of 0.50 means that the two group means differ by 0.50 standard deviations (half a standard deviation). A Cohen's d of 1.20 means that they differ by 1.20 standard deviations. But how should we interpret these values in terms of the strength

of the relationship or the size of the difference between the means? Table 9.4 presents some guidelines for interpreting Cohen's d values in psychological research (Cohen, 1992). Values near 0.20 are considered small, values near 0.50 are considered medium, and values near 0.80 are considered large. Thus a Cohen's d value of 0.50 represents a medium- sized difference between two means, and a Cohen's d value of 1.20 represents a very large difference in the context of social science research. In the research by Ollendick and his colleagues, there was a large difference (d = 0.82) between the exposure and education conditions.

Table 9.3: Strength Evaluation Guidelines for Cohen's d and Pearson's r

Relationship strength	Cohen's d	Pearson's r
Strong/large	0.8	± 0.50
Medium	0.5	± 0.30
Weak/small	0.2	± 0.10

Cohen's d is useful because it has the same meaning regardless of the variable being compared or the scale it was measured on. A Cohen's d of 0.20 means that the two group means differ by 0.20 standard deviations whether we are talking about scores on the Rosenberg Self-Esteem scale, reaction time measured in milliseconds, number of siblings, or diastolic blood pressure measured in millimeters of mercury. Not only does this make it easier for researchers to communicate with each other about their results, it also makes it possible to combine and compare results across different studies using different measures.

Be aware that the term effect size can be misleading because it suggests a causal relationship—that the difference between the two means is an "effect" of being in one group or condition as opposed to another. Imagine, for example, a study showing that a group of exercisers is happier on average than a group of non-exercisers, with an "effect size" of d = 0.35. If the study was an experiment—with participants randomly assigned to exercise and no-exercise conditions—then one could conclude that exercising caused a small to medium-sized increase in happiness. If the study was cross-sectional, however, then one could conclude only that the exercisers were happier than the non-exercisers by a small to medium-sized amount. In other words, simply calling the difference an "effect size" does not make the relationship a causal one.

Correlations Between Quantitative Variables

As we have seen throughout the book, many interesting statistical relationships take the form of correlations between quantitative variables. In general, line graphs are used when the variable on the x-axis has (or is organized into) a small number of distinct values, such as the four quartiles of the name distribution. Scatterplots are used when the variable on the x-axis has a large number of values, such as the different possible self-esteem scores.

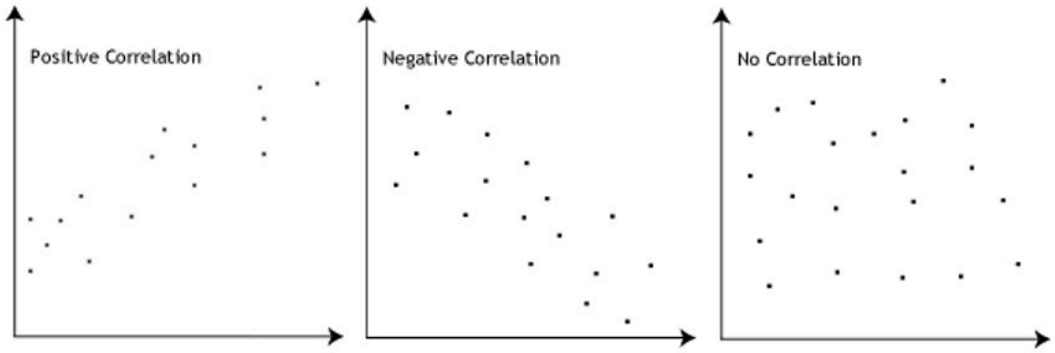


Figure 9.5 Positive, Negative, and No Correlation

The strength of a correlation between quantitative variables is typically measured using a statistic called Pearson's r. Its possible values range from −1.00, through zero, to +1.00. A value of 0 means there is no relationship between the two variables, such as the third image in Figure 9.5. In addition to his guidelines for interpreting Cohen's d, Cohen offered guidelines for interpreting Pearson's r in psychological research (see Table 9.3). Values near ±.10 are considered small, values near ± .30 are considered medium, and values near ±.50 are considered large. Notice that the sign of Pearson's r is unrelated to its strength. Pearson's r values of +.30 and −.30, for example, are equally strong; it is just that one represents a moderate positive relationship and the other a moderate negative relationship. Like Cohen's d, Pearson's r is also referred to as a measure of "effect size" even though the relationship may not be a causal one.

There are two common situations in which the value of Pearson's r can be misleading. One is when the relationship under study is nonlinear. This means that it is important to make a scatterplot and confirm that a relationship is approximately linear before using Pearson's r (or by transforming variables, which is beyond the scope of this text). The other is when one or both of the variables have a limited range in the sample relative to the population. This problem is referred to as restriction of range. Assume, for example, that there is a strong negative correlation between people's age and their enjoyment of hip hop music. However, if we were to collect data only from 18- to 24-year-olds, then the relationship might seem to be quite weak (i.e., enjoyment of hip hop doesn't vary much). In this case, Pearson's r for this restricted range of ages might be zero, even though the relationship holds across all ages. It is a good idea, therefore, to design studies to avoid

restriction of range. For example, if age is one of your primary variables, then you can plan to collect data from people of a wide range of ages. Because restriction of range is not always anticipated or easily avoidable, however, it is good practice to examine your data for possible restriction of range and to interpret Pearson's r in light of it.

Chapter 9: Descriptive Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Expressing Your Results

Once you have conducted your descriptive statistical analyses, you will need to present them to others. In this section, we focus on presenting descriptive statistical results in writing, in figures, and in tables— following American Psychological Association (APA) guidelines for written research reports. These principles can be adapted easily to other presentation formats such as posters and slide show presentations.

Presenting Descriptive Statistics in Writing

Recall that APA style includes several rules for presenting numerical results in the text. These include using words only for numbers less than 10 that do not represent precise statistical results and using numerals for numbers 10 and higher. However, statistical results are always presented in the form of numerals rather than words and are usually rounded to two decimal places (e.g., "2.00" rather than "two" or "2"). They can be presented either in the narrative description of the results or parenthetically—much like reference citations. When you have a small number of results to report, it is often most efficient to write them out. Here are some examples:

- The mean age of the participants was 22.43 years with a standard deviation of 2.34.
- Among the participants with low self-esteem, those in a negative mood expressed stronger intentions to have unprotected sex (M = 4.05, SD = 2.32) than those in a positive mood (M = 2.15, SD = 2.27).
- The treatment group had a mean of 23.40 (SD = 9.33), while the control group had a mean of 20.87 (SD= 8.45).
- The test-retest correlation was .96.
- There was a moderate negative correlation between the alphabetical position of respondents' last names and their response time (r = −.27).

Notice that when presented in the narrative, the terms mean and standard deviation are written out, but when presented parenthetically, the symbols M and SD are used instead. Notice also that it is especially important to use parallel construction to express similar or comparable results in similar ways.

Presenting Descriptive Statistics in Figures

When you have a large number of results to report, you can often do it more clearly and efficiently with a graphical depiction of the data, such as pie charts, bar graphs, or scatterplots. In an APA style research report, these graphs are presented as figures. When you prepare figures for an APA-style research report, there are some general guidelines that you should keep in mind. First, the figure should always add important information rather than repeat information that already appears in the text or in a table (if a figure presents information more clearly or efficiently, then you should keep the figure and eliminate the text or table.) Second, figures should be as simple as possible. For example, the Publication Manual discourages the use of color unless it is absolutely necessary (although color can still be an effective element in posters, slide show presentations, or textbooks.) Third, figures should be interpretable on their own. A reader should be able to understand the basic result based only on the figure and its caption and should not have to refer to the text for an explanation.

There are also several more technical guidelines for presentation of figures that include the following:

- Layout of graphs
- In general, scatterplots, bar graphs, and line graphs should be slightly wider than they are tall.
- The independent variable should be plotted on the x-axis and the dependent variable on the y-axis.
- Values should increase from left to right on the x-axis and from bottom to top on the y-axis.
- The x-axis and y-axis should begin with the value zero.
- Axis Labels and Legends
- Axis labels should be clear and concise and include the units of measurement if they do not appear in the caption.
- Axis labels should be parallel to the axis.
- Legends should appear within the figure.
- Text should be in the same simple font throughout and no smaller than 8 point and no larger than 14 point.
- Captions
- Captions are titled with the word “Figure”, followed by the figure number in the order in which it appears in the text, and terminated with a period. This title is italicized.
- After the title is a brief description of the figure terminated with a period (e.g., “Reaction times of the control versus experimental group.”)
- Following the description, include any information needed to interpret the figure, such as any abbreviations, units of measurement (if not in the axis label), units of error bars, etc.

Bar Graphs

Bar graphs are generally used to present and compare the mean scores for two or more groups or conditions. The bar graph in Figure 12.11 is an APA-style version of Figure 12.4. Notice that it conforms to all the guidelines listed. A new element in Figure 12.11 is the smaller vertical bars that extend both upward and downward from the top of each main bar. These are error bars, and they represent the variability in each group or condition. Although they sometimes extend one standard deviation in each direction, they are more likely to extend one standard error in each direction (as in Figure 9.6). The standard error is the standard deviation of the group divided by the square root of the sample size of the group. The standard error is used because, in general, a difference between group means that is greater than two standard errors is statistically significant. Thus one can "see" whether a difference is statistically significant based on a bar graph with error bars.

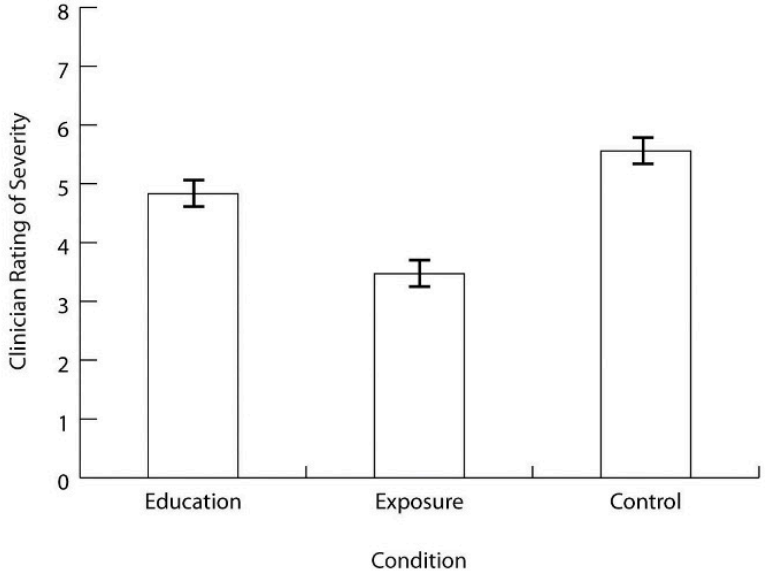


Figure 9.6 Sample APA-Style Bar Graph, With Error Bars Representing the Standard Errors, Based on Research by Ollendick and Colleagues.

Line Graphs

Line graphs are used when the independent variable is measured in a more continuous manner (e.g., time) or to present correlations between quantitative variables when the independent variable has, or is organized into, a relatively small number of distinct levels. Each point in a line graph represents the mean score on the dependent variable for participants at one level of the independent variable. Figure 9.7 is an APA-style version of the results of Carlson and Conard. Notice that it includes error bars representing the standard error and conforms to all the stated guidelines.

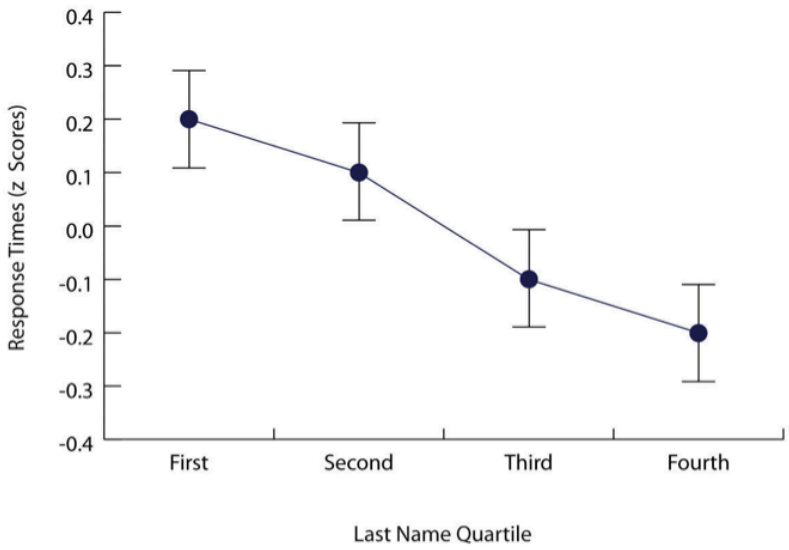


Figure X. Mean response time by the alphabetical position of respondents' names in the alphabet. Response times are expressed as **z** scores. Error bars represent standard errors.

Figure 9.7 Sample APA-Style Line Graph Based on Research by Carlson and Conard.

In most cases, the information in a line graph could just as easily be presented in a bar graph. In Figure 9.7, for example, one could replace each point with a bar that reaches up to the same level and leave the error bars right where they are. This emphasizes the fundamental similarity of the two types of statistical relationship. Both are differences in the average score on one variable across levels of another. The convention followed by most researchers, however, is to use a bar graph when the variable plotted on the x-axis is categorical and a line graph when it is continuous.

Scatterplots

Scatterplots are used to present correlations and relationships between quantitative variables when the variable on the x-axis (typically the independent variable) has a large number of levels. Each point in a scatterplot represents an individual rather than the mean for a group of individuals, and there are no lines connecting the points. The graph in Figure 9.8 is an APA-style scatterplot. Note, when the variables on the x-axis and y-axis are conceptually similar and measured on the same scale—as here, where they are measures of the same variable on two different occasions—this can be emphasized by making the axes the same length. Also, when two or more individuals fall at exactly the same point on the graph, one way this can be indicated is by offsetting the points slightly along the x-axis.

Other ways are by displaying the number of individuals in parentheses next to the point or by making the point larger or darker in proportion to the number of individuals. Finally, the straight line that best fits the points in the scatterplot, which is called the regression line, can also be included.

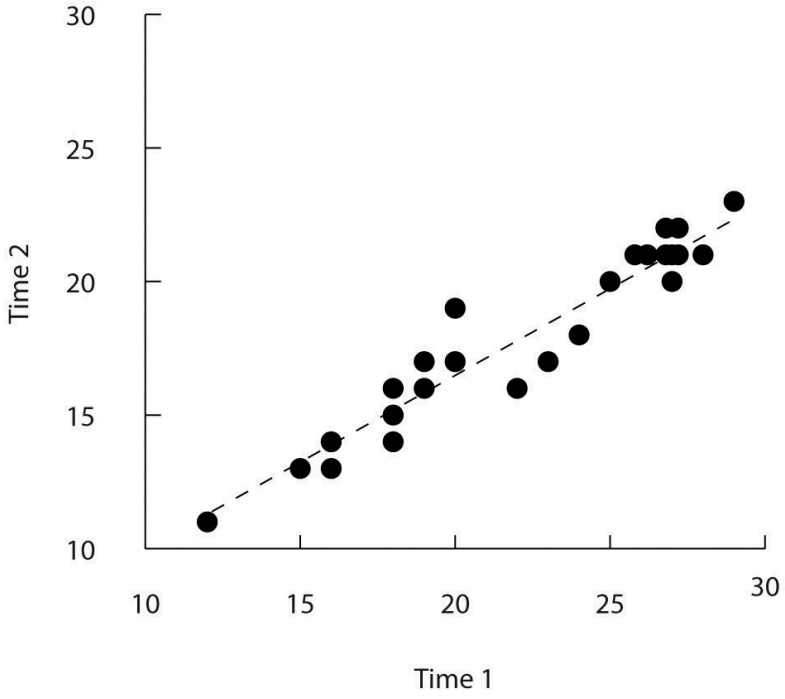


Figure X. Relationship between scores on the Rosenberg self-esteem scale taken by 25 research methods students on two occasions one week apart. Pearson's r =.96.

Figure 9.8 Sample APA-Style Scatterplot

Expressing Descriptive Statistics in Tables

Like graphs, tables can be used to present large amounts of information clearly and efficiently. The same general principles apply to tables as apply to graphs. They should add important information to the presentation of your results, be as simple as possible, and be interpretable on their own. Again, we focus here on tables for an APA-style manuscript.

Figure 9.9, below, reproduced from the Online Writing Lab at Pursue University (OWL, 2022), illustrates a generic table including the following elements:

- **Stub headings** describe the lefthand column, or stub column, which usually lists major independent variables.
- **Column headings** describe entries below them, applying to just one column.
- **Column spanners** are headings that describe entries below them, applying to two or more

columns which each have their own column heading. Column spanners are often stacked on top of column headings and together are called decked heads.

- **Table Spanners** cover the entire width of the table, allowing for more divisions or combining tables with identical column headings. They are the only type of heading that may be plural (OWL, 2022)

Figure 9.9 Sample APA-Style Table

Table 1

Title

Stub Heading	Column Spanner		Column Spanner	
	Column Heading	Column Heading	Column Heading	Column Heading
		Table Spanner		
Row 1	123	234[a]	456	789
Row 2	123	987	543	876
		Table Spanner		
Row 3	432	567	543	908
Row 4	256	849	407*	385

Note. This is a general note, referring to information about the entire table. Notes should be double spaced.

[a]Specific notes appear in a new paragraph; further specific notes follow in the same paragraph.

*A probability note appears in a new paragraph.

As with graphs, precise statistical results that appear in a table do not need to be repeated in the text. Instead, the writer can note major trends and alert the reader to details (e.g., specific correlations) that are of particular interest.

Chapter 9: Descriptive Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology (4th ed.)*** and is licensed under a **CC BY-NC-SA** Licence.

Conducting Your Analyses

Even when you understand the statistics involved, analyzing data can be a complicated process. It is likely that for each of several participants, there are data for several different variables: demographics such as sex and age, one or more independent variables, one or more dependent variables, and perhaps a manipulation check. Furthermore, the "raw" (unanalyzed) data might take several different forms—completed paper-and- pencil questionnaires, computer files filled with numbers or text, videos, or written notes—and these may have to be organized, coded, or combined in some way. There might even be missing, incorrect, or just "suspicious" responses that must be dealt with. In this section, we consider some practical advice to make this process as organized and efficient as possible.

Prepare Your Data for Analysis

Whether your raw data are on paper or in a computer file (or both), there are a few things you should do before you begin analyzing them. First, be sure they do not include any information that might identify individual participants and be sure that you have a secure location where you can store the data and a separate secure location where you can store any consent forms. Unless the data are highly sensitive, a locked room or password-protected computer is usually good enough. It is also a good idea to make photocopies or backup files of your data and store them in yet another secure location—at least until the project is complete. Professional researchers usually keep a copy of their raw data and consent forms for several years in case questions about the procedure, the data, or participant consent arise after the project is completed.

Next, you should check your raw data to make sure that they are complete and appear to have been accurately recorded (whether it was participants, yourself, or a computer program that did the recording). At this point, you might find that there are illegible or missing responses, or obvious misunderstandings (e.g., a response of "12" on a 1-to-10 rating scale). You will have to decide whether such problems are severe enough to make a participant's data unusable. If information about the main independent or dependent variable is missing, or if several responses are missing or suspicious, you may have to exclude that participant's data from the analyses. If you do decide to exclude any data, do not throw them away or delete them because you or another researcher might want to see them later. Instead, set them aside and keep notes about why you decided to exclude them because you will need to report this information.

Now you are ready to enter your data in a spreadsheet program or, if it is already in a computer file, to format it for analysis. You can use a general spreadsheet program like Microsoft Excel or a statistical analysis program like SPSS to create your data file. (Data files created in one program can usually be converted to work with other programs.) The most common format is for each row to represent a participant and for each column to represent a variable (with the variable name at the top of each column). A sample data file is shown in Table 12.6. The first column contains participant identification numbers. This is followed by columns containing demographic information (sex and age), independent variables (mood, four self-esteem items, and the total of the four self-esteem items), and finally dependent variables (intentions and attitudes). Categorical variables can usually be entered as category labels (e.g., "M" and "F" for male and female) or as numbers (e.g., "0" for negative mood and "1" for positive mood). Although category labels are often clearer, some analyses might require numbers. SPSS allows you to enter numbers but also attach a category label to each number.

Table 9.4: Sample Data File

ID	SEX	AGE	MOOD	SE1	SE2	SE3	SE4	TOTAL	INT	ATT
1	M	20	1	2	3	2	3	10	6	5
2	F	22	1	1	0	2	1	4	4	4
3	F	19	0	2	2	2	2	8	2	3
4	F	24	0	3	3	2	3	11	5	6

If you have multiple-response measures—such as the self-esteem measure in Table 9.4—you could combine the items by hand and then enter the total score in your spreadsheet. However, it is much better to enter each response as a separate variable in the spreadsheet and use the software to combine them (e.g., using the "AVERAGE" function in Excel or the "Compute" function in SPSS). Not only is this approach more accurate, but it allows you to detect and correct errors, to assess internal consistency, and to analyze individual responses if you decide to do so later.

Preliminary Analyses

Before turning to your primary research questions, there are often several preliminary analyses to conduct. For multiple-response measures, you should assess the internal consistency of the measure. Statistical programs like SPSS will allow you to compute Cronbach's α or Cohen's κ. If this is beyond your comfort level, you can still compute and evaluate a split-half correlation.

Next, you should analyze each important variable separately. (This step is not necessary for manipulated independent variables, of course, because you as the researcher determined what the distribution would be.) Make histograms for each one, note their shapes, and compute the common measures of central tendency and variability. Be sure you understand what these statistics mean in terms of the variables you are interested in. For example, a distribution of self-report happiness ratings on a 1-to-10-point scale might be unimodal and negatively skewed with a mean of 8.25 and a standard deviation of 1.14. But what this means is that most participants rated themselves fairly high on the happiness scale, with a small number rating themselves noticeably lower.

Now is the time to identify outliers, examine them more closely, and decide what to do about them. You might discover that what at first appears to be an outlier is the result of a response being entered incorrectly in the data file, in which case you only need to correct the data file and move on. Alternatively, you might suspect that an outlier represents some other kind of error, misunderstanding, or lack of effort by a participant. For example, in a reaction time distribution in which most participants took only a few seconds to respond, a participant who took 3 minutes to respond would be an outlier. It seems likely that this participant did not understand the task (or at least was not paying very close attention). Also, including their reaction time would have a large impact on the mean and standard deviation for the sample. In situations like this, it can be justifiable to exclude the outlying response or participant from the analyses. If you do this, however, you should keep notes on which responses or participants you have excluded and why, and apply those same criteria consistently to every response and every participant. When you present your results, you should indicate how many responses or participants you excluded and the specific criteria that you used. And again, do not literally throw away or delete the data that you choose to exclude. Just set them aside because you or another researcher might want to see them later.

Keep in mind that outliers do not necessarily represent an error, misunderstanding, or lack of effort. They might represent truly extreme responses or participants. For example, in one large university student sample, the vast majority of participants reported having had fewer than 15 sexual partners, but there were also a few extreme scores of 60 or 70 (Brown & Sinclair, 1999). Although these scores might represent errors, misunderstandings, or even intentional exaggerations, it is also plausible that they represent honest and even accurate estimates. One strategy here would be to use the median and other statistics that are not strongly affected by the outliers. Another would be to analyze the data both including and excluding any outliers. If the results are essentially the

same, which they often are, then it makes sense to leave the outliers. If the results differ depending on whether the outliers are included or excluded them, then both analyses can be reported and the differences between them discussed.

Understand Your Descriptive Statistics

In the next chapter, we will consider inferential statistics—a set of techniques for deciding whether the results for your sample are likely to apply to the population. Although inferential statistics are important for reasons that will be explained shortly, beginning researchers sometimes forget that their descriptive statistics really tell "what happened" in their study. For example, imagine that a treatment group of 50 participants has a mean score of 34.32 (SD = 10.45), a control group of 50 participants has a mean score of 21.45 (SD = 9.22), and Cohen's d is an extremely strong 1.31. Although conducting and reporting inferential statistics (like a t test) would certainly be a required part of any formal report on this study, it should be clear from the descriptive statistics alone that the treatment worked. Or imagine that a scatterplot shows an indistinct "cloud" of points and Pearson's r is a trivial −.02. Again, although conducting and reporting inferential statistics would be a required part of any formal report on this study, it should be clear from the descriptive statistics alone that the variables are essentially unrelated. The point is that you should always be sure that you thoroughly understand your results at a descriptive level first, and then move on to the inferential statistics.

Chapter 9: Descriptive Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Key Takeaways, Exercises, and References

KEY TAKEAWAYS

- Every variable has a distribution—a way that the scores are distributed across the levels. The distribution can be described using a frequency table and histogram. It can also be described in words in terms of its shape, including whether it is unimodal or bimodal, and whether it is symmetrical or skewed.
- The central tendency, or middle, of a distribution can be described precisely using three statistics—the mean, median, and mode. The mean is the sum of the scores divided by the number of scores, the median is the middle score, and the mode is the most common score.
- The variability, or spread, of a distribution can be described precisely using the range and standard deviation. The range is the difference between the highest and lowest scores, and the standard deviation is the average amount by which the scores differ from the mean.
- The location of a score within its distribution can be described using percentile ranks or z scores. The percentile rank of a score is the percentage of scores below that score, and the z score is the difference between the score and the mean divided by the standard deviation.
- Differences between groups or conditions are typically described in terms of the means and standard deviations of the groups or conditions or in terms of Cohen's d and are presented in bar graphs.
- Cohen's d is a measure of relationship strength (or effect size) for differences between two group or condition means. It is the difference of the means divided by the standard deviation. In general, values of ±0.20, ±0.50, and ±0.80 can be considered small, medium, and large, respectively.
- Correlations between quantitative variables are typically described in terms of Pearson's r and presented in line graphs or scatterplots.
- Pearson's r is a measure of relationship strength (or effect size) for relationships between quantitative variables. It is the mean cross-product of the two sets of z scores. In general, values of ±.10, ±.30, and ±.50 can be considered small, medium, and large, respectively.
- In an APA-style article, simple results are most efficiently presented in the text, while more complex results are most efficiently presented in graphs or tables.
- APA style includes several rules for presenting numerical results in the text. These include using words only for numbers less than 10 that do not represent precise statistical results, and rounding results to two decimal places, using words (e.g., "mean") in the text and symbols (e.g., "M") in parentheses.
- APA style includes several rules for presenting results in graphs and tables. Graphs and tables

should add information rather than repeating information, be as simple as possible, and be interpretable on their own with a descriptive caption (for graphs) or a descriptive title (for tables).
- Raw data must be prepared for analysis by examining them for possible errors, organizing them, and entering them into a spreadsheet program.
- Preliminary analyses on any data set include checking the reliability of measures, evaluating the effectiveness of any manipulations, examining the distributions of individual variables, and identifying outliers.
- Outliers that appear to be the result of an error, a misunderstanding, or a lack of effort can be excluded from the analyses. The criteria for excluded responses or participants should be applied in the same way to all the data and described when you present your results. Excluded data should be set aside and then than destroyed or deleted in case they are needed later.
- Descriptive statistics tell the story of what happened in a study. Although inferential statistics are also important, it is essential to understand the descriptive statistics first.

Exercises

- Make a frequency table and histogram for the following data. Then write a short description of the shape of the distribution in words.
  11, 8, 9, 12, 9, 10, 12, 13, 11, 13, 12, 6, 10, 17, 13, 11, 12, 12, 14, 14
- For the data in Exercise 1, compute the mean, median, mode, standard deviation, and range.
- Using the data in Exercises 1 and 2, find
- the percentile ranks for scores of 9 and 14
- the z scores for scores of 8 and 12.
- The hypothetical data that follow are extraversion scores and the number of Facebook friends for 15 university students. Make a scatterplot for these data, compute Pearson's r, and describe the relationship in words.

Extraversion	Facebook Friends
8	75
10	315
4	28
6	214
12	176
14	95
10	120
11	150
4	32
13	250
5	99
7	136
8	185
11	88
10	144

References

Bem, D. J. (2003) Writing the empirical journal article. In J. M. Darley, M. P. Zanna, & H. L. Roediger III (Eds.), The complete academic: A career guide (2nd ed., pp. 185–219). Washington, DC: American Psychological Association.

Brown, N. R., & Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. The Journal of Sex Research, 36, 292–297.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: A contextual evolutionary analysis of human mating. Psychological Review, 100, 204–232.

Carlson, K. A., & Conard, J. M. (2011). The last name effect: How last name influences acquisition timing. Journal of Consumer Research, 38(2), 300-307. doi: 10.1086/658470

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155–159.

Hyde, J. S. (2007). New directions in the study of gender similarities and differences. Current Directions in Psychological Science, 16, 259–263.

MacDonald, T. K., & Martineau, A. M. (2002). Self-esteem, mood, and intentions to use condoms: When does low self- esteem lead to risky health behaviors? Journal of Experimental Social Psychology, 38, 299–306.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning. Neuropsychology, 24(2), 222–243. doi:10.1037/a0017619

Ollendick, T. H., Öst, L.-G., Reuterskiöld, L., Costa, N., Cederlund, R., Sirbu, C.,...Jarrett, M. A. (2009). One-session treatments of specific phobias in youth: A randomized clinical trial in the United States and Sweden. Journal of Consulting and Clinical Psychology, 77, 504–516.

https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/apa_tables_and_figures.html

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. Journal of Personality and Social Psychology, 89, 623–642.

Chapter 9: Descriptive Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

CHAPTER 10: INFERENTIAL STATISTICS

Matthias Mehl and his colleagues, in their study of sex differences in talkativeness, found that the women in their sample spoke a mean of 16,215 words per day and the men a mean of 15,669 words per day (Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007)1. But despite this sex difference in their sample, they concluded that there was no evidence of a sex difference in talkativeness in the population. Recall also that Allen Kanner and his colleagues, in their study of the relationship between daily hassles and symptoms, found a correlation of +.60 in their sample (Kanner, Coyne, Schaefer, & Lazarus, 1981)2. But they concluded that this finding means there is a relationship between hassles and symptoms in the population. This assertion raises the question of how researchers can say whether their sample result reflects something that is true of the population.

The answer to this question is that they use a set of techniques called inferential statistics, which is what this chapter is about. We focus, in particular, on null hypothesis testing, the most common approach to inferential statistics in psychological research. We begin with a conceptual overview of null hypothesis testing, including its purpose and basic logic. Then we look at several null hypothesis testing techniques for drawing conclusions about differences between means and about correlations between quantitative variables. Finally, we consider a few other important ideas related to null hypothesis testing, including some that can be helpful in planning new studies and interpreting results. We also look at some long-standing criticisms of null hypothesis testing and some ways of dealing with these criticisms.

LEARNING OBJECTIVES

- Understand the logic underpinning hypothesis testing
- Understand the concept of alpha-value and its role in hypothesis testing
- Be familiar with the concept of p-value and its role in hypothesis testing
- Understand under what circumstances it would be appropriate to use the t-test, analysis of variance, correlation analysis, and regression analysis
- Be familiar with the role of "statistical power" in hypothesis testing
- Be aware of the common criticisms of null hypothesis testing

Chapter 10: Inferential Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Understanding Null Hypothesis Testing

As we have seen, psychological research typically involves measuring one or more variables in a sample and computing descriptive summary data (e.g., means, correlation coefficients) for those variables. These descriptive data for the sample are called statistics. In general, however, the researcher's goal is not to draw conclusions about that sample but to draw conclusions about the population that the sample was selected from. Thus researchers must use sample statistics to draw conclusions about the corresponding values in the population. These corresponding values in the population are called parameters. Imagine, for example, that a researcher measures the number of depressive symptoms exhibited by each of 50 adults with clinical depression and computes the mean number of symptoms. The researcher probably wants to use this sample statistic (the mean number of symptoms for the sample) to draw conclusions about the corresponding population parameter (the mean number of symptoms for adults with clinical depression).

Unfortunately, sample statistics are not perfect estimates of their corresponding population parameters. This is because there is a certain amount of random variability in any statistic from sample to sample. The mean number of depressive symptoms might be 8.73 in one sample of adults with clinical depression, 6.45 in a second sample, and 9.44 in a third—even though these samples are selected randomly from the same population. Similarly, the correlation (Pearson's r) between two variables might be +.24 in one sample, −.04 in a second sample, and +.15 in a third—again, even though these samples are selected randomly from the same population. This random variability in a statistic from sample to sample is called sampling error. (Note that the term error here refers to random variability and does not imply that anyone has made a mistake. No one "commits a sampling error.")

One implication of this is that when there is a statistical relationship in a sample, it is not always clear that there is a statistical relationship in the population. A small difference between two group means in a sample might indicate that there is a small difference between the two group means in the population. But it could also be that there is no difference between the means in the population and that the difference in the sample is just a matter of sampling error. Similarly, a Pearson's r value of −.29 in a sample might mean that there is a negative relationship in the population. But it could also be that there is no relationship in the population and that the relationship in the sample is just a matter of sampling error.

In fact, any statistical relationship in a sample can be interpreted in two ways:

- There is a relationship in the population, and the relationship in the sample reflects this.
- There is no relationship in the population, and the relationship in the sample reflects only sampling error.

The purpose of null hypothesis testing is simply to help researchers decide between these two interpretations.

The Logic of Null Hypothesis Testing

Null hypothesis testing (often called null hypothesis significance testing or NHST) is a formal approach to deciding between two interpretations of a statistical relationship in a sample. One interpretation is called the null hypothesis (often symbolized H0 and read as "H-zero"). This is the idea that there is no relationship in the population and that the relationship in the sample reflects only sampling error. Informally, the null hypothesis is that the sample relationship "occurred by chance." The other interpretation is called the alternative hypothesis (often symbolized as H1). This is the idea that there is a relationship in the population and that the relationship in the sample reflects this relationship in the population.

Again, every statistical relationship in a sample can be interpreted in either of these two ways: It might have occurred by chance, or it might reflect a relationship in the population. So researchers need a way to decide between them. Although there are many specific null hypothesis testing techniques, they are all based on the same general logic. The steps are as follows:

- Assume for the moment that the null hypothesis is true. There is no relationship between the variables in the population.
- Determine how likely the sample relationship would be if the null hypothesis were true.
- If the sample relationship would be extremely unlikely, then reject the null hypothesis in favor of the alternative hypothesis. If it would not be extremely unlikely, then retain the null hypothesis.

Following this logic, we can begin to understand why Mehl and his colleagues concluded that there is no difference in talkativeness between women and men in the population. In essence, they asked the following question: "If there were no difference in the population, how likely is it that we would find a small difference of d = 0.06 in our sample?" Their answer to this question was that this sample relationship would be fairly likely if the null hypothesis were true. Therefore, they retained the null hypothesis—concluding that there is no evidence of a sex difference in the population. We can also see why Kanner and his colleagues concluded that there is a correlation between hassles and symptoms in the population. They asked, "If the null hypothesis were true, how likely is it that we would find a strong correlation of +.60 in our sample?" Their answer to this question was that this sample relationship would be fairly unlikely if the null hypothesis were true. Therefore, they rejected the null hypothesis in favor of the alternative hypothesis—concluding that there is a positive correlation between these variables in the population.

A crucial step in null hypothesis testing is finding the probability of the sample result or a more extreme result if the null hypothesis were true (Lakens, 2017).1 This probability is called the p value. A low p value means that the sample or more extreme result would be unlikely if the null hypothesis were true and leads to the rejection of the null hypothesis. A p value that is not low means that the sample or more extreme result would be likely if the null hypothesis were true and leads to the retention of the null hypothesis. But how low must the p value criterion be before the sample result is considered unlikely enough to reject the null hypothesis? In null hypothesis testing, this criterion is called α (alpha) and is almost always set to .05. If there is a 5% chance or less of a result at least as extreme as the sample result if the null hypothesis were true, then the null hypothesis is rejected. When this happens, the result is said to be statistically significant. If there is greater than a 5% chance of a result as extreme as the sample result when the null hypothesis is true, then the null hypothesis is retained. This does not necessarily mean that the researcher accepts the null hypothesis as true— only that there is not currently enough evidence to reject it. Researchers often use the expression "fail to reject the null hypothesis" rather than "retain the null hypothesis," but they never use the expression "accept the null hypothesis."

Chapter 10: Inferential Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Role of Sample Size and Relationship Strength

Recall that null hypothesis testing involves answering the question, "If the null hypothesis were true, what is the probability of a sample result as extreme as this one?" In other words, "What is the p value?" It can be helpful to see that the answer to this question depends on just two considerations: the strength of the relationship and the size of the sample. Specifically, the stronger the sample relationship and the larger the sample, the less likely the result would be if the null hypothesis were true. That is, the lower the p value. This should make sense. Imagine a study in which a sample of 500 women is compared with a sample of 500 men in terms of some psychological characteristic, and Cohen's d is a strong 0.50. If there were really no sex difference in the population, then a result this strong based on such a large sample should seem highly unlikely. Now imagine a similar study in which a sample of three women is compared with a sample of three men, and Cohen's d is a weak 0.10. If there were no sex difference in the population, then a relationship this weak based on such a small sample should seem likely. And this is precisely why the null hypothesis would be rejected in the first example and retained in the second.

Of course, sometimes the result can be weak and the sample large, or the result can be strong and the sample small. In these cases, the two considerations trade off against each other so that a weak result can be statistically significant if the sample is large enough and a strong relationship can be statistically significant even if the sample is small. Table 13.1 shows roughly how relationship strength and sample size combine to determine whether a sample result is statistically significant. The columns of the table represent the three levels of relationship strength: weak, medium, and strong. The rows represent four sample sizes that can be considered small, medium, large, and extra large in the context of psychological research. Thus each cell in the table represents a combination of relationship strength and sample size. If a cell contains the word Yes, then this combination would be statistically significant for both Cohen's d and Pearson's r. If it contains the word No, then it would not be statistically significant for either. There is one cell where the decision for d and r would be different and another where it might be different depending on some additional considerations, which are discussed in the section "Some Basic Null Hypothesis Tests."

Table 10.1 Relationship Strength, Sample Size and Statistical Significance

Table 10.1: Relationship Strength, Sample Size and Statical Significance

	Relationship strength		
Sample Size	**Weak**	**Medium**	**Strong**
Small (N = 20)	No	No	d = Maybe r = Yes
Medium (N = 50)	No	Yes	Yes
Large (N = 100)	d = Yes r = No	Yes	Yes
Extra large (N = 500)	Yes	Yes	Yes

Although Table 10.1 provides only a rough guideline, it shows very clearly that weak relationships based on medium or small samples are never statistically significant and that strong relationships based on medium or larger samples are always statistically significant. If you keep this lesson in mind, you will often know whether a result is statistically significant based on the descriptive statistics alone. It is extremely useful to be able to develop this kind of intuitive judgment. One reason is that it allows you to develop expectations about how your formal null hypothesis tests are going to come out, which in turn allows you to detect problems in your analyses. For example, if your sample relationship is strong and your sample is medium, then you would expect to reject the null hypothesis. If for some reason your formal null hypothesis test indicates otherwise, then you need to double-check your computations and interpretations. A second reason is that the ability to make this kind of intuitive judgment is an indication that you understand the basic logic of this approach in addition to being able to do the computations.

Statistical Significance versus Practical Significance

Table 10.1 illustrates another extremely important point. A statistically significant result is not necessarily a strong one. Even a very weak result can be statistically significant if it is based on a large enough sample. This is closely related to Janet Shibley Hyde's argument about sex differences (Hyde, 2007). The differences between women and men in mathematical problem solving and leadership ability are statistically significant.

But the word significant can cause people to interpret these differences as strong and important—perhaps even important enough to influence the university courses they take or even who they vote for. As we have seen, however, these statistically significant differences are actually quite weak—perhaps even "trivial."

This is why it is important to distinguish between the statistical significance of a result and the practical significance of that result. Practical significance refers to the importance or usefulness of the result in some real-world context. Many sex differences are statistically significant—and may even be interesting for purely scientific reasons—but they are not practically significant. In clinical practice, this same concept is often referred to as "clinical significance." For example, a study on a new treatment for social phobia might show that it produces a statistically significant positive effect. Yet this effect still might not be strong enough to justify the time, effort, and other costs of putting it into practice—especially if easier and cheaper treatments that work almost as well already exist. Although statistically significant, this result would be said to lack practical or clinical significance.

Chapter 10: Inferential Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Some Basic Null Hypothesis Tests

The t-Test

As we have seen throughout this book, many studies in psychology focus on the difference between two means. The most common null hypothesis test for this type of statistical relationship is the t-test. In this section, we look at three types of t tests that are used for slightly different research designs: the one- sample t-test, the dependent-samples t- test, and the independent-samples t-test. You may have already taken a course in statistics, but we will refresh your statistical knowledge.

One-Sample t-Test

The one-sample t-test is used to compare a sample mean (M) with a hypothetical population mean ($\mu 0$) that provides some interesting standard of comparison. The null hypothesis is that the mean for the population (μ) is equal to the hypothetical population mean: $\mu = \mu 0$. The alternative hypothesis is that the mean for the population is different from the hypothetical population mean: $\mu \neq \mu 0$. To decide between these two hypotheses, we need to find the probability of obtaining the sample mean (or one more extreme) if the null hypothesis were true. But finding this p value requires first computing a test statistic called t. (A test statistic is a statistic that is computed only to help find the p value.)

The reason the t statistic (or any test statistic) is useful is that we know how it is distributed when the null hypothesis is true. In this case, the distribution is unimodal and symmetrical, and it has a mean of 0. Its precise shape depends on a statistical concept called the degrees of freedom, which for a one-sample t-test is $N - 1$. The important point is that knowing this distribution makes it possible to find the p value for any t score.

Fortunately, we do not have to deal directly with the distribution of t scores. If we were to enter our sample data and hypothetical mean of interest into one of the online statistical tools or into a program like SPSS (Excel does not have a one-sample t-test function), the output would include both the t score and the p value. At this point, the rest of the procedure is simple. If p is equal to or less than .05, we reject the null hypothesis and conclude that the population mean differs from the hypothetical mean of interest. If p is greater than .05, we retain the null hypothesis and con-

clude that there is not enough evidence to say that the population mean differs from the hypothetical mean of interest. (Again, technically, we conclude only that we do not have enough evidence to conclude that it does differ.)

Thus far, we have considered what is called a two-tailed test, where we reject the null hypothesis if the t score for the sample is extreme in either direction. This test makes sense when we believe that the sample mean might differ from the hypothetical population mean but we do not have good reason to expect the difference to go in a particular direction. But it is also possible to do a one-tailed test, where we reject the null hypothesis only if the t score for the sample is extreme in one direction that we specify before collecting the data. This test makes sense when we have good reason to expect the sample mean will differ from the hypothetical population mean in a particular direction.

Here is how it works: We simply redefine extreme to refer only to one tail of the distribution. If 5% of the values of t beyond the critical value for t are all in one tail of the distribution, the advantage of the one-tailed test is that critical values are less extreme. If the sample mean differs from the hypothetical population mean in the expected direction, then we have a better chance of rejecting the null hypothesis. The disadvantage is that if the sample mean differs from the hypothetical population mean in the unexpected direction, then there is no chance at all of rejecting the null hypothesis.

Example: One-Sample t–Test

Imagine that a health psychologist is interested in the accuracy of university students' estimates of the number of calories in a chocolate chip cookie. He shows the cookie to a sample of 10 students and asks each one to estimate the number of calories in it. Because the actual number of calories in the cookie is 250, this is the hypothetical population mean of interest ($\mu 0$). The null hypothesis is that the mean estimate for the population (μ) is 250. Because he has no real sense of whether the students will underestimate or overestimate the number of calories, he decides to do a two-tailed test. Now imagine further that the participants' actual estimates are as follows:

250, 280, 200, 150, 175, 200, 200, 220, 180, 250.

The mean estimate for the sample (M) is 212.00 calories and the standard deviation (SD) is 39.17. The health psychologist can now compute the t score for his sample. If he enters the data into one of the online analysis tools or uses SPSS, it would tell him that the two- tailed p value for the computed t score (with $10 - 1 = 9$ degrees of freedom) is .013. Because this is less than .05, the health psychologist would reject the null hypothesis and conclude that university students tend to underestimate the number of calories in a chocolate chip cookie.

Finally, if this researcher had gone into this study with good reason to expect that university students underestimate the number of calories, then he could have done a one-tailed test instead of a two-tailed test. The only thing this decision would change is the critical value, which would be

−1.833. This slightly less extreme value would make it a bit easier to reject the null hypothesis. However, if it turned out that university students overestimate the number of calories—no matter how much they overestimate it—the researcher would not have been able to reject the null hypothesis.

The Dependent-Samples t–Test

The dependent-samples t-test (sometimes called the paired-samples t-test) is used to compare two means for the same sample tested at two different times or under two different conditions. This comparison is appropriate for pretest-posttest designs or within-subjects experiments. The null hypothesis is that the means at the two times or under the two conditions are the same in the population. The alternative hypothesis is that they are not the same. This test can also be one-tailed if the researcher has good reason to expect the difference goes in a particular direction.

It helps to think of the dependent-samples t-test as a special case of the one-sample t-test. However, the first step in the dependent-samples t-test is to reduce the two scores for each participant to a single difference score by taking the difference between them. At this point, the dependent-samples t-test becomes a one-sample t-test on the difference scores. The hypothetical population mean ($\mu 0$) of interest is 0 because this is what the mean difference score would be if there were no difference on average between the two times or two conditions. We can now think of the null hypothesis as being that the mean difference score in the population is 0 ($\mu 0 = 0$) and the alternative hypothesis as being that the mean difference score in the population is not 0 ($\mu 0 \neq 0$).

Example: Dependent-Samples t–Test

Imagine that the health psychologist now knows that people tend to underestimate the number of calories in junk food and has developed a short training program to improve their estimates. To test the effectiveness of this program, s/he conducts a pretest-posttest study in which 10 participants estimate the number of calories in a chocolate chip cookie before the training program and then again afterward. Because s/he expects the program to increase the participants' estimates, s/he decides to do a one-tailed test. Now imagine further that the pretest estimates are:

230, 250, 280, 175, 150, 200, 180, 210, 220, 190

and that the posttest estimates (for the same participants in the same order) are:

250, 260, 250, 200, 160, 200, 200, 180, 230, 240.

The difference scores, then, are as follows:

20, 10, −30, 25, 10, 0, 20, −30, 10, 50.

Note that it does not matter whether the first set of scores is subtracted from the second or the second from the first as long as it is done the same way for all participants. In this example, it makes sense to subtract the pretest estimates from the posttest estimates so that positive difference scores mean that the estimates went up after the training and negative difference scores mean the estimates went down.

If s/he enters the data into one of the online analysis tools or uses Excel or SPSS, it would output that the one-tailed p value for this t score (again with 10 − 1 = 9 degrees of freedom) is .148. Because this is greater than .05, s/he would retain the null hypothesis and conclude that the training program does not significantly increase people's calorie estimates.

The Independent-Samples t-Test

The independent-samples t-test is used to compare the means of two separate samples (M1 and M2). The two samples might have been tested under different conditions in a between-subjects experiment, or they could be pre-existing groups in a cross-sectional design (e.g., women and men, extraverts and introverts). The null hypothesis is that the means of the two populations are the same: $\mu 1 = \mu 2$. The alternative hypothesis is that they are not the same: $\mu 1 \neq \mu 2$. Again, the test can be one-tailed if the researcher has good reason to expect the difference goes in a particular direction.

Example: Independent-Samples t–Test

Now the health psychologist wants to compare the calorie estimates of people who regularly eat junk food with the estimates of people who rarely eat junk food. S/he believes the difference could come out in either direction so s/he decides to conduct a two-tailed test. S/he collects data from a sample of eight participants who eat junk food regularly and seven participants who rarely eat junk food. The data are as follows:

- Junk food eaters: 180, 220, 150, 85, 200, 170, 150, 190
- Non–junk food eaters: 200, 240, 190, 175, 200, 300, 240

If s/he enters the data into one of the online analysis tools or uses Excel or SPSS, it would indicate that the two- tailed p value for this t score (with 15 − 2 = 13 degrees of freedom) is .015. Because this p value is less than .05, the health psychologist would reject the null hypothesis and conclude that people who eat junk food regularly make lower calorie estimates than people who eat it rarely.

The Analysis of Variance

T-tests are used to compare two means (a sample mean with a population mean, the means of two conditions or two groups). When there are more than two groups or condition means to be compared, the most common null hypothesis test is the analysis of variance (ANOVA). In this section, we look primarily at the one-way ANOVA, which is used for between-subjects designs with a single independent variable.

One-Way ANOVA

The one-way ANOVA is used to compare the means of more than two samples (M1, M2...MG) in a between-subjects design. The null hypothesis is that all the means are equal in the population: $\mu 1 = \mu 2 = \ldots = \mu G$. The alternative hypothesis is that not all the means in the population are equal.

The test statistic for the ANOVA is called F. The reason that F is useful is that we know how it is distributed when the null hypothesis is true. This distribution is unimodal and positively skewed with values that cluster around 1. The precise shape of the distribution depends on both the number of groups and the sample size, and there are degrees of freedom values associated with each of these. The between-groups degrees of freedom is the number of groups minus one: $dfB = (G - 1)$. The within-groups degrees of freedom is the total sample size minus the number of groups: $dfW = N - G$. Again, knowing the distribution of F when the null hypothesis is true allows us to find the p value.

Statistical software such as Excel and SPSS will compute F and find the p value. If p is equal to or less than .05, then we reject the null hypothesis and conclude that there are differences among the group means in the population. If p is greater than .05, then we retain the null hypothesis and conclude that there is not enough evidence to say that there are differences.

Example: One-Way ANOVA

Imagine that a health psychologist wants to compare the calorie estimates of psychology majors, nutrition majors, and professional dieticians. He collects the following data:

- Psych majors: 200, 180, 220, 160, 150, 200, 190, 200
- Nutrition majors: 190, 220, 200, 230, 160, 150, 200, 210, 195
- Dieticians: 220, 250, 240, 275, 250, 230, 200, 240

The means are 187.50 (SD = 23.14), 195.00 (SD = 27.77), and 238.13 (SD = 22.35), respectively. So it appears that dieticians made substantially more accurate estimates on average. The researcher would almost certainly enter these data into a program such as Excel or SPSS, which would compute F for him or her and find the p value. Table 13.4 shows the output of the one-way ANOVA function in Excel for these data. This table is referred to as an ANOVA table. It shows that MSB is 5,971.88, MSW is 602.23, and their ratio, F, is 9.92. The p value is .0009. Because this value is below .05, the researcher would reject the null hypothesis and conclude that the mean calorie estimates for the three groups are not the same in the population.

Table 10.2: Typical One-Way ANOVA Output from Excel

Source of variation	SS	df	MS	F	p-value	Fcrit
Between groups	11,943.75	2	5,971.88	9.916234	0.000928	3.4668
Within groups	12,646.88	21	602.2321			
Total	24,590.63	23				

ANOVA Elaborations Post Hoc Comparisons

When we reject the null hypothesis in a one-way ANOVA, we conclude that the group means are not all the same in the population. But this can indicate different things. With three groups, it can indicate that all three means are significantly different from each other, or it can indicate that one of the means is significantly different from the other two, but the other two are not significantly different from each other. It could be, for example, that the mean calorie estimates of psychology majors, nutrition majors, and dieticians are all significantly different from each other. Or it could be that the mean for dieticians is significantly different from the means for psychology and nutrition majors, but the means for psychology and nutrition majors are not significantly different from each other. For this reason, statistically significant one-way ANOVA results are typically followed up with a series of post hoc comparisons of selected pairs of group means to determine which are different from which others.

One approach to post hoc comparisons would be to conduct a series of independent-samples t-tests comparing each group mean to each of the other group means. But there is a problem with this approach. In general, if we conduct a t-test when the null hypothesis is true, we have a 5% chance of mistakenly rejecting the null hypothesis (see Section 13.3 “Additional Considerations” for more on such Type I errors). If we conduct several t-tests when the null hypothesis is true, the chance of mistakenly rejecting at least one null hypothesis increases with each test we conduct.

Thus researchers do not usually make post hoc comparisons using standard t-tests because there is too great a chance that they will mistakenly reject at least one null hypothesis. Instead, they use one of several modified t-test procedures—among them the Bonferonni procedure, Fisher's least significant difference (LSD) test, and Tukey's honestly significant difference (HSD) test. The details of these approaches are beyond the scope of this book, but it is important to understand their purpose. It is to keep the risk of mistakenly rejecting a true null hypothesis to an acceptable level (close to 5%).

Testing Correlation Coefficients

For relationships between quantitative variables, where Pearson's r (the correlation coefficient) is used to describe the strength of those relationships, the appropriate null hypothesis test is a test of the correlation coefficient. The basic logic is exactly the same as for other null hypothesis tests. In this case, the null hypothesis is that there is no relationship in the population. We can use the Greek lowercase rho (ρ) to represent the relevant parameter: $\rho = 0$. The alternative hypothesis is that there is a relationship in the population: $\rho \neq 0$. As with the t- test, this test can be two-tailed if the researcher has no expectation about the direction of the relationship or one-tailed if the researcher expects the relationship to go in a particular direction.

It is possible to use the correlation coefficient for the sample to compute a t score with $N - 2$ degrees of freedom and then to proceed as for a t-test. However, because of the way it is computed, the correlation coefficient can also be treated as its own test statistic. The online statistical tools and statistical software such as Excel and SPSS generally compute the correlation coefficient and provide the p value associated with that value. As always, if the p value is equal to or less than .05, we reject the null hypothesis and conclude that there is a relationship between the variables in the population. If the p value is greater than .05, we retain the null hypothesis and conclude that there is not enough evidence to say there is a relationship in the population.

Example: Test of a Correlation Coefficient

Imagine that the health psychologist is interested in the correlation between people's calorie estimates and their weight. She has no expectation about the direction of the relationship, so she decides to conduct a two- tailed test. She computes the correlation coefficient for a sample of 22 university students and finds that Pearson's r is −.21. The statistical software she uses tells her that the p value is .348. It is greater than .05, so she retains the null hypothesis and concludes that there is no relationship between people's calorie estimates and their weight.

Simple Regression

Regression is a special kind of correlation analysis, but in the case of regression it isn't the strength of the association that one is interested in but rather how a change in one variable may be accompanied by a change in another variable. For instance, what effect on average does studying 15 minutes more each day have on a student's GPA? As such, regression is a predictive exercise—we want to predict how a dependent variable will change given a particular change in the independent variable.

In order to estimate these effects, we need to derive a regression equation, which is simply an equation that describes the relationship between the two variables. The general form of a simple regression equation is: y = mx + b. Y is the dependent (outcome) variable, x is the independent variable, m is the slope of the line that describes the relationship between them, and b is a constant that indicates where the line crosses the y axis.

For example, consider the following:

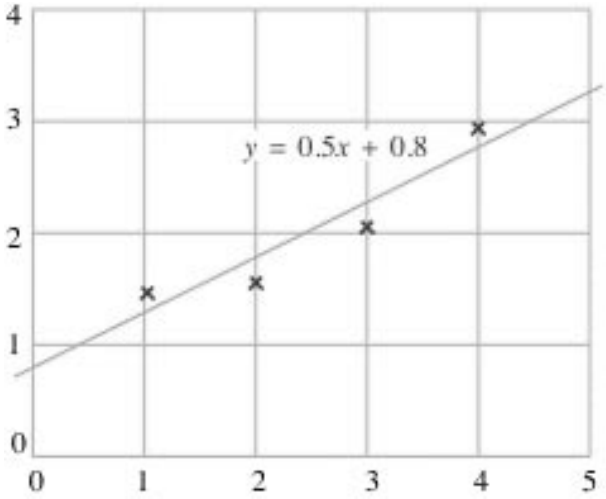


Figure 10.1
Regression Analysis Example

In this case, our dependent variable is predicted by the equation y = 0.5x + 0.8. Let's suppose that variable x is body satisfaction and variable y is self esteem. If one's body satisfaction were 5, we would expect one's self esteem to be 0.5(5) + 0.8 = 3.3.

The way that regression equations are most often used, however, is not for this kind of specific prediction (at least, not in the way just described) but rather they are used to gauge the effect of a change in the independent variable upon the dependent variable. For example, what effect might a one unit increase in body satisfaction have upon one's level of self esteem? Using the equation above, we would say that a one unit increase in body satisfaction would lead to a 0.5(1) + 0.8 = 1.3 unit increase in self esteem.

Multiple Regression

That which has just been described is called simple regression because there is only one independent variable. Multiple regression works the same way but it is used when you have more than one independent variable. For instance, perhaps you have developed a model (what researchers call equations like these) that has perceived health being a function of exercise frequency, weight, and perceived healthfulness of foods typically consumed. In this case, the formula would be something like: PH = βEF + βW + βHF + b, where each of the "β"s represents a particular coefficient that is applied to each independent variable (like 0.5 was in the previous example). These "betas" tell you the kind of effect a change in one independent variable will have *while all the other variables remain constant*. So, the bigger the beta coefficient, the bigger the effect upon the dependent variable will be. Suppose the equation were actually PH = 5EF + 2 W + 8HF. A one unit increase in perceived healthfulness of food consumed (HF) will yield an eight unit increase in perceived health, holding the other variables constant. Similarly, a one unit increase in weight (W) should yield a 2 unit increase in PH, and a 1 unit increase in exercise frequency (EF) should yield a 5 unit increase in PH. So, if you wanted to make people feel as though they were healthier, which personal characteristic would you spend your time trying to increase (perhaps through some kind of counselling or therapy)? Based on this equation, if you could get someone to increase their perception that they eat healthy foods, this would have the biggest effect on their overall health perceptions.

One small disclaimer need be said: That's not exactly how regression works. It is actually based on what are called "standardized values," but explaining what these are and how they are derived would a) take too long and b) not really add a heck of a lot to the explanation, anyway.

In short, standardizing scores transforms them all into the same unit of measure (actually, standard deviation units) so that they can reasonably be compared; in essence, it turns apples and oranges into apples and apples. That way, when one talks about a "one unit increase" this means the same thing irrespective of whether one variable is measured on a scale from one to four, another in pounds, and another in miles.

Chapter 10: Inferential Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Additional Considerations

In this section, we consider a few other issues related to null hypothesis testing, including some that are useful in planning studies and interpreting results. We even consider some long-standing criticisms of null hypothesis testing, along with some steps that researchers in psychology have taken to address them.

Errors in Null Hypothesis Testing

In null hypothesis testing, the researcher tries to draw a reasonable conclusion about the population based on the sample. Unfortunately, this conclusion is not guaranteed to be correct. This discrepancy is illustrated by Figure 10.3. The rows of this table represent the two possible decisions that researchers can make in null hypothesis testing: to reject or retain the null hypothesis. The columns represent the two possible states of the world: the null hypothesis is false or it is true. The four cells of the table, then, represent the four distinct outcomes of a null hypothesis test. Two of the outcomes—rejecting the null hypothesis when it is false and retaining it when it is true—are correct decisions. The other two—rejecting the null hypothesis when it is true and retaining it when it is false—are errors.

Table 10.3: Errors in Null Hypothesis Testing

	True state of the world	
Decision	**H_0=true**	**H_0=false**
Reject H_0	Type 1 error	Correct
DNR H_0	Correct	Type 2 error

Rejecting the null hypothesis when it is true is called a Type I error. This error means that we have concluded that there is a relationship in the population when in fact there is not. Type I errors occur because even when there is no relationship in the population, sampling error alone will occasionally produce an extreme result. In fact, when the null hypothesis is true and α is .05, we will mistakenly reject the null hypothesis 5% of the time. (This possibility is why α is sometimes referred to as the "Type I error rate.") Retaining the null hypothesis when it is false is called a Type II error.

This error means that we have concluded that there is no relationship in the population when in fact there is a relationship. In practice, Type II errors occur primarily because the research design lacks adequate statistical power to detect the relationship (e.g., the sample is too small). We will have more to say about statistical power shortly.

In principle, it is possible to reduce the chance of a Type I error by setting α to something less than .05. Setting it to .01, for example, would mean that if the null hypothesis is true, then there is only a 1% chance of mistakenly rejecting it. But making it harder to reject true null hypotheses also makes it harder to reject false ones and therefore increases the chance of a Type II error. Similarly, it is possible to reduce the chance of a Type II error by setting α to something greater than .05 (e.g., .10). But making it easier to reject false null hypotheses also makes it easier to reject true ones and therefore increases the chance of a Type I error. This provides some insight into why the convention is to set α to .05. There is some agreement among researchers that the .05 level of α keeps the rates of both Type I and Type II errors at acceptable levels.

The possibility of committing Type I and Type II errors has several important implications for interpreting the results of our own and others' research. One is that we should be cautious about interpreting the results of any individual study because there is a chance that it reflects a Type I or Type II error. This possibility is why researchers consider it important to replicate their studies. Each time researchers replicate a study and find a similar result, they rightly become more confident that the result represents a real phenomenon and not just a Type I or Type II error.

Statistical Power

The statistical power of a research design is the probability of rejecting the null hypothesis given the sample size and expected relationship strength. For example, the statistical power of a study with 50 participants and an expected Pearson's r of +.30 in the population is .59. That is, there is a 59% chance of rejecting the null hypothesis if indeed the population correlation is +.30. Statistical power is the complement of the probability of committing a Type II error. So in this example, the probability of committing a Type II error would be 1 − .59 = .41. Clearly, researchers should be interested in the power of their research designs if they want to avoid making Type II errors. In particular, they should make sure their research design has adequate power before collecting data. A common guideline is that a power of .80 is adequate. This guideline means that there is an 80% chance of rejecting the null hypothesis for the expected relationship strength.

The topic of how to compute power for various research designs and null hypothesis tests is beyond the scope of this book. However, there are online tools that allow you to do this by entering your sample size, expected relationship strength, and α level for various hypothesis tests (see "Computing Power Online"). In addition, Table 10.4 shows the sample size needed to achieve a power of .80 for weak, medium, and strong relationships for a two-tailed independent-samples t-test and for

a two-tailed test of Pearson's r. Notice that this table amplifies the point made earlier about relationship strength, sample size, and statistical significance. In particular, weak relationships require very large samples to provide adequate statistical power.

Table 10.4 Sample Sizes Needed to Achieve Statistical Power of .80

	Null Hypothesis Test	
Relationship Strength	**Independent-Samples t-Test**	**Test of Pearson's r**
Strong (d = .80, r = .50)	52	28
Medium (d = .50, r = .30)	128	84
Weak (d = .20, r = .10)	788	782

What should you do if you discover that your research design does not have adequate power? Imagine, for example, that you are conducting a between-subjects experiment with 20 participants in each of two conditions and that you expect a medium difference (d = .50) in the population. The statistical power of this design is only .34. That is, even if there is a medium difference in the population, there is only about a one in three chance of rejecting the null hypothesis and about a two in three chance of committing a Type II error. Given the time and effort involved in conducting the study, this probably seems like an unacceptably low chance of rejecting the null hypothesis and an unacceptably high chance of committing a Type II error.

Given that statistical power depends primarily on relationship strength and sample size, there are essentially two steps you can take to increase statistical power: increase the strength of the relationship or increase the sample size. Increasing the strength of the relationship can sometimes be accomplished by using a stronger manipulation or by more carefully controlling extraneous variables to reduce the amount of noise in the data (e.g., by using a within-subjects design rather than a between-subjects design). The usual strategy, however, is to increase the sample size. For any expected relationship strength, there will always be some sample large enough to achieve adequate power.

Chapter 10: Inferential Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Criticisms of Null Hypothesis Testing

Again, null hypothesis testing is the most common approach to inferential statistics. It is not without its critics, however. Some criticisms of null hypothesis testing focus on researchers' misunderstanding of it. We have already seen, for example, that the p value is widely misinterpreted as the probability that the null hypothesis is true. (Recall that it is really the probability of the sample result if the null hypothesis were true.) A closely related misinterpretation is that 1 – p equals the probability of replicating a statistically significant result. In one study, 60% of a sample of professional researchers thought that a p value of .01—for an independent- samples t-test with 20 participants in each sample—meant there was a 99% chance of replicating the statistically significant result (Oakes, 1986)4. Our earlier discussion of power should make it clear that this figure is far too optimistic. As Table 13.5 shows, even if there were a large difference between means in the population, it would require 26 participants per sample to achieve a power of .80. And the program G*Power shows that it would require 59 participants per sample to achieve a power of .99.

Another set of criticisms focuses on the logic of null hypothesis testing. To many, the strict convention of rejecting the null hypothesis when p is less than .05 and retaining it when p is greater than .05 makes little sense. This criticism does not have to do with the specific value of .05 but with the idea that there should be any rigid dividing line between results that are considered significant and results that are not. Imagine two studies on the same statistical relationship with similar sample sizes. One has a p value of .04 and the other a p value of .06. Although the two studies have produced essentially the same result, the former is likely to be considered interesting and worthy of publication and the latter simply not significant. This convention is likely to prevent good research from being published and to contribute to the file drawer problem.

Yet another set of criticisms focus on the idea that null hypothesis testing—even when understood and carried out correctly—is simply not very informative. Recall that the null hypothesis is that there is no relationship between variables in the population (e.g., Cohen's d or Pearson's r is precisely 0). So to reject the null hypothesis is simply to say that there is some nonzero relationship in the population. But this assertion is not really saying very much. Imagine if chemistry could tell us only that there is some relationship between the temperature of a gas and its volume—as opposed to providing a precise equation to describe that relationship. Some critics even argue that the relationship between two variables in the population is never precisely 0 if it is carried out to enough decimal places. In other words, the null hypothesis is never literally true. So rejecting it does not tell us anything we did not already know!

To be fair, many researchers have come to the defense of null hypothesis testing. One of them, Robert Abelson, has argued that when it is correctly understood and carried out, null hypothesis testing does serve an important purpose (Abelson, 1995). Especially when dealing with new phenomena, it gives researchers a principled way to convince others that their results should not be dismissed as mere chance occurrences.

Chapter 10: Inferential Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

What to do?

Even those who defend null hypothesis testing recognize many of the problems with it. But what should be done? Some suggestions now appear in the APA Publication Manual. One is that each null hypothesis test should be accompanied by an effect size measure such as Cohen's d or Pearson's r. By doing so, the researcher provides an estimate of how strong the relationship in the population is—not just whether there is one or not. (Remember that the p value cannot substitute as a measure of relationship strength because it also depends on the sample size. Even a very weak result can be statistically significant if the sample is large enough.)

Another suggestion is to use confidence intervals rather than null hypothesis tests. A confidence interval around a statistic is a range of values that is computed in such a way that some percentage of the time (usually 95%) the population parameter will lie within that range. For example, a sample of 20 university students might have a mean calorie estimate for a chocolate chip cookie of 200 with a 95% confidence interval of 160 to 240. In other words, there is a very good (95%) chance that the mean calorie estimate for the population of university students lies between 160 and 240. Advocates of confidence intervals argue that they are much easier to interpret than null hypothesis tests. Another advantage of confidence intervals is that they provide the information necessary to do null hypothesis tests should anyone want to. In this example, the sample mean of 200 is significantly different at the .05 level from any hypothetical population mean that lies outside the confidence interval. So the confidence interval of 160 to 240 tells us that the sample mean is statistically significantly different from a hypothetical population mean of 250 (because the confidence interval does not include the value of 250).

Finally, there are more radical solutions to the problems of null hypothesis testing that involve using very different approaches to inferential statistics. Bayesian statistics, for example, is an approach in which the researcher specifies the probability that the null hypothesis and any important alternative hypotheses are true before conducting the study, conducts the study, and then updates the probabilities based on the data. It is too early to say whether this approach will become common in social scientific research. For now, null hypothesis testing—supported by effect size measures and confidence intervals—remains the dominant approach.

Chapter 10: Inferential Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology (4th ed.)*** and is licensed under a **CC BY-NC-SA** Licence.

Key Takeaways, Exercises, and References

Key Takeaways

- Null hypothesis testing is a formal approach to deciding whether a statistical relationship in a sample reflects a real relationship in the population or is just due to chance.
- The logic of null hypothesis testing involves assuming that the null hypothesis is true, finding how likely the sample result would be if this assumption were correct, and then making a decision. If the sample result would be unlikely if the null hypothesis were true, then it is rejected in favor of the alternative hypothesis. If it would not be unlikely, then the null hypothesis is retained.
- The probability of obtaining the sample result if the null hypothesis were true (the p value) is based on two considerations: relationship strength and sample size. Reasonable judgments about whether a sample relationship is statistically significant can often be made by quickly considering these two factors.
- Statistical significance is not the same as relationship strength or importance. Even weak relationships can be statistically significant if the sample size is large enough. It is important to consider relationship strength and the practical significance of a result in addition to its statistical significance.
- To compare two means, the most common null hypothesis test is the t- test. The one-sample t-test is used for comparing one sample mean with a hypothetical population mean of interest, the dependent- samples t-test is used to compare two means in a within-subjects design, and the independent- samples t-test is used to compare two means in a between-subjects design.
- To compare more than two means, the most common null hypothesis test is the analysis of variance (ANOVA). The one-way ANOVA is used for between-subjects designs with one independent variable, the repeated- measures ANOVA is used for within-subjects designs, and the factorial ANOVA is used for factorial designs.
- A null hypothesis test of Pearson's r is used to compare a sample value of Pearson's r with a hypothetical population value of 0.
- Regression analysis seeks to understand relationships between variables via an algebraic expression which describes the relationship and how changes in the value of independent variable/s affect the value of the outcome variable.
- The decision to reject or retain the null hypothesis is not guaranteed to be correct. A Type I error occurs when one rejects the null hypothesis when it is true. A Type II error occurs when one fails to reject the null hypothesis when it is false.

- The statistical power of a research design is the probability of rejecting the null hypothesis given the expected strength of the relationship in the population and the sample size. Researchers should make sure that their studies have adequate statistical power before conducting them.
- Null hypothesis testing has been criticized on the grounds that researchers misunderstand it, that it is illogical, and that it is uninformative. Others argue that it serves an important purpose—especially when used with effect size measures, confidence intervals, and other techniques. It remains the dominant approach to inferential statistics in the social sciences.

Exercises

- Imagine a study showing that people who eat more broccoli tend to be happier. Explain for someone who knows nothing about statistics why the researchers would conduct a null hypothesis test.
- A sample of 25 university students rated their friendliness on a scale of 1 (Much Lower Than Average) to 7 (Much Higher Than Average). Their mean rating was 5.30 with a standard deviation of 1.50. Conduct a one-sample t-test comparing their mean rating with a hypothetical mean rating of 4 (Average). The question is whether university students have a tendency to rate themselves as friendlier than average.
- A researcher compares the effectiveness of two forms of psychotherapy for social phobia using an independent-samples t-test.
- Explain what it would mean for the researcher to commit a Type I error.
- Explain what it would mean for the researcher to commit a Type II error.
- Imagine that you conduct a t-test and the p value is .02. How could you explain what this p value means to someone who is not already familiar with null hypothesis testing? Be sure to avoid the common misinterpretations of the p value.

References

Abelson, R. P. (1995). Statistics as principled argument. Mahwah, NJ: Erlbaum.

Cohen, J. (1994). The world is round: p < .05. American Psychologist, 49, 997–1003.

Hyde, J. S. (2007). New directions in the study of gender similarities and differences. Current Directions in Psychological Science, 16, 259–263.

Kanner, A. D., Coyne, J. C., Schaefer, C., & Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. Journal of Behavioral Medicine, 4, 1–39.

Lakens, D. (2017, December 25). About p-values: Understanding common misconceptions. [Blog post] Retrieved from https://correlaid.org/en/blog/understand-p-values/

Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? Science, 317, 82.

Oakes, M. (1986). Statistical inference: A commentary for the social and behavioral sciences. Chichester, UK: Wiley.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. Psychological Bulletin, 83, 638–641.

Simonsohn U., Nelson L. D., & Simmons J. P. (2014). P-Curve: a key to the file drawer. Journal of Experimental Psychology: General, 143(2), 534–547. doi: 10.1037/a0033242

Tramimow, D. & Marks, M. (2015). Editorial. Basic and Applied Social Psychology, 37, 1–2. https://dx.doi.org/10.1080/ 01973533.2015.1012991

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594–604.

Chapter 10: Inferential Statistics is adapted from Rajiv S. Jhangiani, Carrie Cuttler, and Dana C. Leighton (2019) ***Research Methods in Psychology* (4th ed.)** and is licensed under a **CC BY-NC-SA** Licence.

Versioning History

PETER MORDEN

This page provides a record of edits and changes made to this book since its initial publication. If the change is minor, the version number increases by 0.1. If the edits involve substantial updates, the version number increases to the next full number. Due to the nature of the open textbook being continuously updated, the addition or removal of a resource is not recorded on this page.

For feedback get in touch with Peter Morden

Version	Date	Change	Affected Web Page
1.0		Remix and revision of Principles of Sociological Inquiry: Qualitative and Quantitative Methods, Chapter 1-8, Cuttler et al.'s Research Methods in Psychology (4th ed.), Chapter 12-13, and De Carlo's Scientific Inquiry in Social Work, Chapter 12. Removal of images, restructuring of chapters for flow and clarity Revision for appropriateness to students in Recreation & Leisure Studies, Therapeutic Recreation, and Human Relations	All