2

Research Methods and Analysis

Big Questions	Learning Objectives
1. What research methods are used to study relationships?	**2.1** Describe the scientific method, different types of research study, and ethical considerations.
2. How are results analyzed?	**2.2** Compare and contrast the most common ways to analyze and interpret research results.
3. What is "open science"?	**2.3** Analyze the "open science" movement and what it means for future research endeavors.

The scientific study of attraction, love, friendship, and everything else covered in this book is a hugely impressive feat.

We're talking about squishy, abstract concepts that even the people experiencing them can't explain. So anyone who decides to jump into the deep end of the science pool by trying to define, measure, and experiment on the world of intimate relationships is up for a challenge! And as you read about the theories and findings covered in this book, it's always important to ask *how* scientists arrived at these claims. Think of methods and statistics as the building blocks to finding answers, or a treasure map leading toward the riches of insight. A fundamental knowledge about different research approaches and how results are analyzed is essential to a true understanding of the science of relationships.

What Research Methods Are Used to Study Relationships?

As we've already started to discuss, the scientific study of love, attraction, friendship, sexuality, romance, and all the behaviors, thoughts, and emotions involved is a difficult task. Even the question of defining or operationalizing variables such as "love" is done in a widely different number of ways across different studies, as you read in Chapter 1. Let's start by considering how relationships researchers ask and answer questions from a scientific approach. Then, we'll review five different approaches to setting up a study. For an interesting link on how pioneering sex researchers have been portrayed in TV and movies, see the "Relationships and Popular Culture" feature as well.

The Scientific Method

Most academic disciplines that study intimate relationships approach the topic from the **scientific method**, a systematic and evidence-based approach to asking and answering questions. The general approach of the scientific method is displayed in Figure 2.1. Researchers will start by observing a pattern in the "real world" and will then generate a formal **hypothesis**, or a specific statement of what they believe will happen in a study designed to test the phenomenon of interest.

RELATIONSHIPS AND POPULAR CULTURE

Scientifically Studying Sex: Pioneers on the Screen

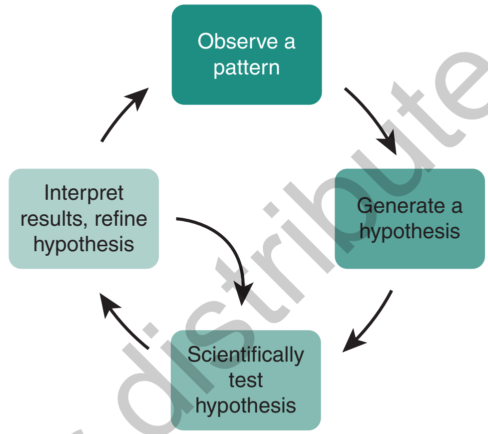
In the first half of the 1900s, studying human sexual behavior was quite controversial—even scandalous. One of the most controversial figures in the history of research on sex and intimate relationships was, unquestionably, Alfred Kinsey. One part of his legacy might even be the terms "sexology" and "sexologist," as he developed university courses on human sexuality and created one of the largest research efforts to understand the true nature of human sexual behaviors. Part of Kinsey's controversy were his views that women were capable of several different kinds of orgasm (e.g., from both vaginal and clitoral stimulation) and that everyone is at least a little bit bisexual. However, as shocking as those ideas might have been in the 1930s–1950s when he was a prominent professor, his research methodology itself was also controversial. Kinsey interviewed prostitutes, prisoners, abuse victims, and gay men—populations that had previously been ignored. He also crossed the boundaries between objective observer and participant as he engaged in sexual behaviors with his participants, graduate students, and colleagues on his research team. In 2004, Kinsey's fascinating life was made into a movie simply called *Kinsey*, starring Liam Neeson and Laura Linney (Coppola, Mutrux, & Condon, 2004).

Following in Kinsey's footsteps were the famous pair Masters and Johnson (specifically, William Masters and Virginia Johnson), who published research on human sexual behaviors over the entire second half of the twentieth century. They also studied phases of sexuality and tried to understand the female orgasm (which is apparently quite the mystery!), including how and why women are able to have multiple orgasms in a short period of time. Their lives and research have been fictionalized in the television series *Masters of Sex* that ran from 2013 to 2016 and starred Michael Sheen and Lizzy Caplan (Ashford, 2013–2016). As you can probably expect, both *Kinsey* and *Masters of Sex* are filled with content that might not be suitable for children to watch.

After operationalizing all the variables involved, making sure the study is ethical, and gathering data, researchers then interpret the results (often using statistical analysis). Once the pattern of results is known, the process can repeat again and again, each time providing a deeper or more detailed understanding of the topic. It's also important to think about standards of quality when designing a study.

There are several ways to analyze whether any given study is done well. First, a study should have solid **internal validity**. Internal validity relates to how well the study was constructed and whether any results can really be interpreted in the ways the researcher intends. For example, were all the variables operationalized and measured appropriately? If there are two groups being compared, are they identical in every way except for the main variable of interest, or are there other explanations at play? Internal validity can be compared to **external validity**, which is also important. While internal validity refers to the structure within the study and whether it was set up correctly, external validity relates to whether the study can be applied to other people or situations. Can the results generalize to people beyond those who actually participated? Does the study have any "real-world" implications?

FIGURE 2.1 ● The Scientific Method



The scientific method starts by noticing interesting patterns and generating hypotheses regarding those patterns. Then, evidence is gathered that either supports or refutes the hypothesis. Those results help us refine hypotheses and keep testing them as we learn more.

Another concern that's receiving more and more attention is the ability to find a **replication** of any study's results. Replication means doing the same study again, with different people, and confirming the results by finding the same patterns over and over. If a study's results can never be found again, it calls into question whether the original study was really done properly. Replication—or lack of replication—can lead to controversy. A famous example occurred in the 1950s. A man named James Vicary claimed that when he hid subliminal messages of popcorn and Coke in film reels at movie theaters, there was a 58% increase in sales of popcorn and an 18% increase in sales of Coca-Cola (Pratkanis, 1992). After several years of scientists attempting to replicate similar behavioral effects of subliminal messages—and failing to do so—Vicary finally had to admit that he had made the entire study up!

This is an extreme example. Failure to replicate usually doesn't mean that the original study was bogus or that the findings aren't interesting and important. It does mean that we should ask questions, such as why other people can't seem to find the same results—and sometimes there are theoretically interesting answers that lead to additional hypotheses. Perhaps there was something special about the original participants, or there was something happening in the world that affected the results, or the sample of people in the study was too small, or the statistics were done incorrectly. . . . But if the researchers were, indeed, honest about their results (as the vast majority of scientists are), a lack of replication might actually lead to interesting developments in theory or practice.

Methodological Approaches

Imagine that you want to design a research study. For the purposes of the next few pages, let's use an example: You're interested in investigating whether introverts or extroverts are happier with the levels of intimacy within their social relationships. (By the way, both "extrovert" and "extravert" are acceptable spellings of the word.) Maybe extroverts (people who are more gregarious and social on most occasions) perceive that they have more close friends in terms of *quantity*, but introverts feel that their relationships are more intimate (e.g., better *quality*). You might make that your hypothesis. As we've already covered, you have to formally state your hypothesis and operationalize your variables (define them and decide how to measure them).

You also have a few other decisions to make, right off the bat. One is whether you will do a qualitative or a quantitative study. A **qualitative study** is one that gathers open-ended data, usually through surveys or interviews, in non-numerical form. You might set up interviews with 10 people who self-identify as introverts and 10 who self-identify as extroverts and ask them questions such as, "Tell me how you feel about the levels of intimacy in your relationships." This straightforward approach has the advantage of allowing the participants to be the experts in their own lives and provides an interesting level of detail and a personal touch in the results you gather.

A real example of a qualitative study is one by Rosen (1996). She interviewed 22 women who were survivors of abusive relationships. She asked the participants why they had initially been attracted to the person who later became their abuser and how they felt once the abuse started. Her interviews led to fascinating insights regarding the mindset of abuse victims (which are described for you in detail in Chapter 12). These women's stories, emotions, and insights might have been lost if they had only been asked questions in the form of numbers, like scales from 1 to 10.

However, you might prefer a **quantitative study**, in which the data you gather are in numerical form and thus better suited for most statistical analyses. Paralleling the theoretical example from before, you might ask the same 10 introverts and 10 extroverts to take surveys in which they answer a series of questions on a scale from 1 (strongly disagree) to 7 (strongly agree), then you average their answers. Quantitative data are useful for gaining an understanding of patterns of results across more people and have the advantage of additional statistical analyses, but they lose the personal feel of qualitative data.

An example of a real quantitative study is a survey used by Arnocky and Vaillancourt (2014). They asked participants to fill out several numerical scales regarding whether victims of relationship abuse were "responsible" for what happened to them (a form of victim blaming). Participants read different scenarios, and the results showed that participants blamed victims more (scores on the quantitative scale were higher) when the scenarios described a male victim, compared to a female victim. In other words, victim blaming appeared to be worse for male victims in this study, a finding based on numerical results that might have been harder to understand if participants only described their feelings.

Another decision to make is whether the study should be cross-sectional or longitudinal. A **cross-sectional study** occurs at a single time, whereas a longitudinal study takes place over two or more time periods or data-collection sessions.

"Cross-sectional" studies are called that because the general idea is that you can compare results across multiple groups at a single time. In contrast, remember the Alameda County Study from Chapter 1 (Berkman & Syme, 1979), which was longitudinal because it followed the participants over many years. Longitudinal studies have the advantage of seeing how patterns change over time. This is especially interesting in a relationships context, to see how friends or couple members change and adapt as their relationship grows in intimacy (or ends!). The disadvantages of longitudinal studies are the time, expense, and effort that they take, plus the fact that many of the participants might drop out of the study before it's done. Because of these disadvantages, cross-sectional studies are more common.

When recruiting participants for a study, it's important to get as many people as possible. In addition, ideally the participants have enough diversity that they represent a wide variety of cultures, sexualities, ages, religions, experiences, and so on. This ideal type of sample is called a **representative sample**, meaning the participants in your study serve as examples of a typical person in the larger population, and that the participants are diverse enough to cover many different perspectives. The best strategy to get a representative sample is to use **random sampling** of the larger population, meaning that everyone in the larger group has an equal chance of participating. That way, your study isn't biased toward only one type of person.

The ideal situation is for theories and hypotheses to be tested many times, with many different participants, in many settings. We can be more and more confident of claims when we replicate studies that have strong internal and external validity. Multiple methods, such as both quantitative and qualitative data, also help. So there are many different ways to test a hypothesis. Most studies used for this book fall into one of five basic structures or types of research methodology; each one is covered next.

Option 1: Archival Research

Archival data are stored pieces of information that were originally created for some other purpose. Newspapers, census data, Facebook posts, and even popular culture are all examples of archival data. To explore our study regarding introvert and extrovert differences in relationship intimacy, you could examine Facebook posts and profiles to see how many "friends" people have, how many times other people post to their page and vice versa, how many "likes" their posts get, how many times the owner of the profile self-discloses intimate information on their page, and so on.

Archival research has led to important insights in the world of intimate relationships. One interesting example is an understanding of abusive marriages. For years, some researchers believed that abusive relationships almost always had male perpetrators and female victims and that violent incidents were fairly severe. This perspective came from looking at archival data collected through police reports and profiles of victims in emergency shelters. However, researchers who were doing studies on abusive marriages through anonymous surveys found very different patterns, including female-to-male violence and many couples in which physically aggressive behavior was mutual (see Johnson, 1995, 2007). This debate is discussed further in Chapter 12, but for now the point is that the source of archival data is important in how it shapes our understanding.

Option 2: Naturalistic Observation

Another approach is **naturalistic observation**, or scientific surveillance of people in their natural environments—in other words, where the behaviors would be occurring anyway, even if you weren't there watching it happen. You might decide to go to a local bar to watch people interact with their friends or flirt with strangers and make notes of the patterns you see play out.

You might be thinking, "If some scientist came to the bar and started writing down everything I did, then I probably wouldn't react very naturally." You're right, and this is a potential challenge to good observational research. When people change their behavior due to awareness that they're being observed, it's called **reactivity**. One creative solution is a technique called **participant observation**, in which scientists disguise themselves as people who belong in that environment. It's kind of like going undercover. You pretend you're not doing research at all and hope to fade into the background—and still find some discreet way to record your observations. In our example, perhaps you pretend to simply be an innocent bar patron, or you act as the bartender so you have an excuse to talk with people about their thoughts.

Participant observation may create some ethical problems, so be careful. After all, you are deceiving people about why you are there. And it may be an ethical violation to observe people when they don't know they are being observed. The advantage of this technique—or any form of naturalistic observation—is that hopefully we get to see authentic social behaviors.

Option 3: Self-Report Surveys

Perhaps the most popular research method in studying intimate relationships is **self-report surveys**, in which people are directly asked to write down their own thoughts, feelings, and behaviors. The types of measures used in self-report surveys are scales like the Rubin liking and loving scales or the scale measuring Sternberg's love components that you saw in Chapter 1.

Pictorial Press Ltd / Alamy Stock Photo

Atlaspix / Alamy Stock Photo

In the movie *21 Jump Street* (Moritz, Cannell, Lord, & Miller, 2012), two young police officers go undercover pretending to be high school students so they can bust a new drug that's hitting the community. In *Imperium* (Taufique, Lee, Ragussis, Walker, & Ragussis, 2016), Daniel Radcliffe's character works for the FBI and infiltrates a White supremacist group, pretending to be racist. If any of them had been social psychologists doing research with this "undercover" technique, it would have been called participant observation.

The main advantages to using self-report scales are that it's relatively inexpensive and fast to get a lot of data, lots of participants can take the survey (making the sample more diverse), and statistical techniques can analyze patterns of responses (if the surveys were quantitative). Self-report surveys are also often the only way to get access to people's intimate personal lives, such as their sexual fantasies, whether they want to cheat on their partner, or what they find sexually attractive.

Remember that one common problem with naturalistic observation is reactivity, or people changing their behaviors because they know they are being observed. Self-report surveys have their own concerns, and one of the big ones is dishonesty. People might not want to admit to cheating, abuse, and other behaviors generally not socially acceptable. The dishonesty problem is called **social desirability**, the idea that people shape their responses—exaggerate, manipulate, or just straight out lie—so that others will have positive impressions of them. (This problem is also sometimes known as impression management.)

WHAT'S MY SCORE?

Measuring Social Desirability

Instructions: Listed below are several statements concerning personal attitudes and traits. Please read each item and decide whether the statement is true or false as it pertains to you, personally. Circle T for true statements and F for false statements.

T F 1. Before voting I thoroughly investigate the qualifications of all the candidates.

T F 2. I never hesitate to go out of my way to help someone in trouble.

T F 3. I sometimes feel resentful when I don't get my way.

T F 4. I am always careful about my manner of dress.

T F 5. My table manners at home are as good as when I eat out in a restaurant.

T F 6. I like to gossip at times.

T F 7. I can remember "playing sick" to get out of something.

T F 8. There have been occasions when I took advantage of someone.

T F 9. I'm always willing to admit it when I make a mistake.

T F 10. There have been occasions when I felt like smashing things.

T F 11. I am always courteous, even to people who are disagreeable.

T F 12. At times I have really insisted on having things my own way.

Scoring: Give yourself 1 point each if you said TRUE for question 1, 2, 4, 5, 9, or 11. Then, give yourself 1 point each if you said FALSE for question 3, 6, 7, 8, 10, or 12. Then, add your points. Higher scores indicate great attempts to manage your impression on others, or a higher tendency toward socially desirable responding on self-report scales.

Source: Crowne and Marlowe (1960).

Critical Thinking: If a participant shows a high level of deception based on this scale, is the only option to ignore the rest of their data in any given research study? All you know is that they might not have been honest—you can't tell in what direction or to what degree they've been dishonest in the rest of their responses. So, what can you do with the rest of their data?

One creative way around this potential problem is to include a measure of social desirability in the survey, to see whether people admit to behaviors that almost everyone does. If someone denies something like gossiping or littering, for example, they're probably not being particularly honest just because pretty much everyone does these things at least occasionally. An example social desirability scale is shown in the "What's My Score?" feature.

Option 4: Quasi-Experiments

Many studies on intimate relationships are interested in comparing two or more groups of people to each other. For example, in our theoretical study, we want to compare introverts to extroverts. But we don't get to manipulate people's personality; we have no control over whether they are introverted or extroverted. When researchers gather data in which two or more *naturally occurring groups* are compared to each other, it's called a **quasi-experiment**. It's "quasi," meaning "half-formed" or "almost" because it's not a "true" experiment (those are explained in the next section).

Another example of a quasi-experiment would be to compare people who are in long-distance relationships to those who are not, to see if the quality of the two different types of relationship changes. When one of my students and I did this (Butler & Goodfriend, 2015), we found that relationship satisfaction levels were similar in each relationship type, but that people *believed* that satisfaction was lower in an "average" long-distance relationship. This belief might eventually lead to a self-fulfilling prophecy of lower satisfaction in long-distance couples—although we didn't test that possibility with longitudinal data. Quasi-experiments are very commonly done because researchers want to compare naturally occurring groups such as Republicans versus Democrats, married versus divorced people, heterosexual versus gay/lesbian relationships, people who grew up in abusive homes versus healthy homes, and so on. These are interesting and important questions, and quasi-experiments are the only way to find answers.

Option 5: Experiments

An **experiment** compares two or more groups of participants who have been formed through **random assignment**. Random assignment means that each participant is put in one of the groups by chance.

Imagine you wanted to know whether listening to love songs made people feel more positively toward their partner. So, you randomly assigned 50 people to listen to love songs (this would be called the **experimental group**) and randomly assigned 50 different people to hear no songs at all (if a "neutral" or comparison group exists in an experiment, it is called the **control group**). If it's true that the two groups really are equal in every way *except* for hearing love songs or not and then the two groups have different outcomes in terms of their feelings toward their partner, then it's fairly safe to conclude that the only possible explanation for their different feelings was the songs. You can say that the love songs caused an increase in feelings of love.

In an experiment, what makes the groups different, based on that random assignment, is called the **independent variable**. The independent variable is what the researchers set up to make Group 1 versus Group 2 (or Group 3, and so on). In the case of the theoretical experiment in the previous paragraph, the independent

TABLE 2.1 ● Examples of Studies, Independent Variables, and Dependent Variables

Experiment Basics	Independent Variable	Dependent Variable
People listen to love songs or no songs and then rate how much they love their partner.	Presence or absence of love songs	Love ratings for partner
People list either positive or negative memories about their partner, then estimate their chances of being married in 10 years.	Memory type (positive or negative)	Estimates of marriage probability
Children play with either white-skinned or brown-skinned dolls, then answer questions about which magazine models they think are the prettiest.	Doll type (white-skinned or brown-skinned)	Perceptions of the models' attractiveness
Two strangers sit in a room together for 1 hour and are asked to get to know each other. The room is either well-lit (lights on) or dark (lights off). Researchers code how intimate their conversation becomes.	Lights on or off in the room	Level of intimacy in conversation

Experiments have independent variables, which separate participants into different groups. They also have dependent variables, or the outcome being measured.

variable is the presence or absence of love songs. There doesn't have to be a control or neutral group, but there always has to be some kind of comparison group. For example, this study could have compared people who listened to love songs with people who listened to jazz, country, or classical music. In a perfect study, the independent variable is the *only* difference between or among groups.

The outcome variable in an experiment is called the **dependent variable**. It's called "dependent" because if the hypothesis is correct, then scores or levels of this variable are "dependent" upon which group the participant was in. Here, the dependent variable is feelings toward a partner, and they're expected to be more positive after love songs are played. So love feelings are "dependent" on whether they heard the songs or not. In short, independent variables are the "cause" in an experiment, and dependent variables are the "effect" or outcome. For several more examples of independent and dependent variables in theoretical studies on intimate relationships, see Table 2.1.

Ethical Considerations Within Research Studies

Any research study done with living creatures needs to be done ethically. Investigations into people's personal, intimate lives is a context in which ethical considerations must be taken very seriously. Consider experiments in which researchers are actually trying to *manipulate* people's feelings, thoughts, or behaviors within their intimate relationships! Or surveys asking people about sexual abuse or domestic violence or secret affairs . . . just answering questions about these topics could lead to traumatic memories and emotions. Research on intimate relationships is challenging both for methodological reasons and for ethical and moral reasons.

There's a certain level of trust that happens when anyone shows up to participate in a research study. As researchers, we want to remember that we have a solemn responsibility to treat people with respect. Even when we use unobtrusive methodologies like naturalistic observation or archival studies, all people involved in the study of human social behavior should be valued. Researchers across all the sciences provide ethical and legal guidance about what it means to treat study participants with respect through **institutional review boards (IRBs)**, which are committees of people who consider the ethical implications of any study. Your local IRB committee is typically composed of representatives from different departments in a college, university, research institute, or corporation. The committees also often have a lawyer as a member, and sometimes they have a member with no background in research at all, to represent the "average person's" perspective.

Some of the participant rights required by most IRBs are the following:

- *Informed Consent:* Participants should be told what they will be asked to do and whether there are any potential dangers or risks involved in the study before it begins; this is called **informed consent**.
- *Deception:* Participants should be told the truth about the purpose and nature of the study as much as possible. **Deception**, or hiding the true nature of the study, is allowed only when it is necessary because knowing the truth would change how the participants respond.
- *Anonymity:* No participant's individual responses will be published in a way that identifies them publicly. Any identifying information (such as names or specific demographics) needs to be removed if individual responses are to be reported, such as answers during a qualitative interview.
- *Right to Withdraw:* Participants have the right to stop being in the study at any time, for any reason, or to skip questions on a survey if they are not comfortable answering them.
- *Debriefing:* After completing the study, all participants should be given additional details about the hypotheses of the study, allowed the opportunity to ask questions, and even see the results if they wish. This **debriefing** after the study is complete should definitely include an explanation of any deception that was involved (if deception occurred) so that participants have the opportunity to withdraw their data if they are upset about the deception.

CHECK YOUR UNDERSTANDING

2.1 Dr. X asks participants to complete a single survey in which they write essay responses to several questions. Everyone answers the same questions. This study is therefore an example of which type of research?

a. Longitudinal
b. Experimental
c. Archival
d. Qualitative

2.2 True experiments use random assignment to place each participant into one of the groups being studied. Random assignment helps ensure the groups are identical in every other way, which means that use of random assignment helps improve the study's:

a. Ethics
b. Replication
c. External validity
d. Internal validity

2.3 Dr. Z asks half of their participants to fill out Rubin's liking and loving scales in a very cold room and the other half to fill out the same scales in a very hot room. What is the dependent variable in this study?

a. The number of participants Dr. Z has in the study
b. Temperature of the room (hot or cold)
c. Participants' answers to the Rubin scales
d. Whether the participants are told Dr. Z's hypothesis

APPLICATION ACTIVITY

Choose a topic you think is very personal or controversial within the field of intimate relationships and search for a study done on that topic. Then, analyze the methodology that was used and discuss whether you think the participants were treated ethically. Were all of the typical IRB standards used, as far as you can tell? If not, do you think the study violated ethical guidelines for research studies? Could the study have been done differently, to improve ethical treatment?

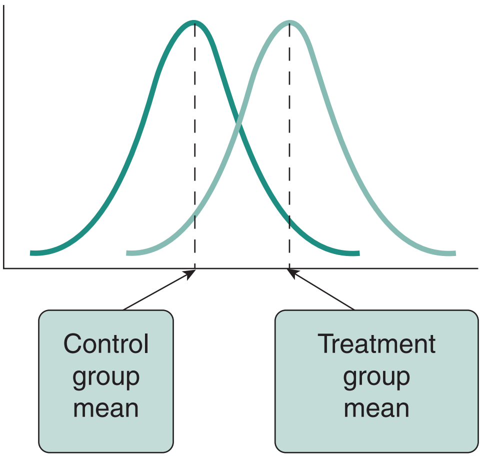
CRITICAL THINKING

- If you wanted to study patterns of attraction or relationship behaviors within the students of your college or university, what kinds of archival data could you use? Consider public resources such as Facebook, library archives, the school newspaper, and so on.
- This section introduced you to a measure of social desirability that could be used to identify if participants are being honest in their responses. Imagine you did a survey study and 20 of 60 participants had high scores on this scale, indicating dishonesty. What do you do now with the rest of their survey answers? Do you throw them out completely? Why or why not? Do you need to collect more data, from other people, so you have more "honest" answers?
- Most people who are monitored during studies using naturalistic observation are never told that they are being watched or that their behavior might end up in a study (even if it is anonymous). Does this practice seem ethical to you? Why or why not?

Answers to the Check Your Understanding Questions

2.1 d, **2.2** d, and **2.3** c.

FIGURE 2.2 ● Comparing Two Groups of Participants: A *t*-Test



One way researchers look for patterns is by comparing average scores between different groups of participants. When we compare two groups, as here, we use a *t*-test. When we compare three or more groups, we do an analysis of variance, or ANOVA.

How Are Results Analyzed?

If you've chosen a quantitative design, you'll need to understand the results of your study using some basic statistical tests. Even if you, personally, never conduct a study, this book contains summaries of many other people's studies, so to understand them you need a foundational understanding of how scientists decide what their data really mean. Let's cover just the basics of two different sets of statistical approach that you'll see throughout this book: comparing groups to each other and doing correlations.

Comparing Groups: *t*-Tests and ANOVAs

In a survey, quasi-experiment, or true experiment, researchers often want to compare answers or behaviors across groups of participants. There are two basic statistical tests we use to do that. The first is called a ***t*-test**, which compares responses between two different groups. It might be men versus women, introverts versus extroverts, happy versus unhappy couples, people who listen to love songs or no songs, or any other two relevant groups of people.

You can remember that a *t*-test compares two groups by thinking that the "*t*" stands for "two." Just as you can see in Figure 2.2, a *t*-test will compare two things in each group: the average of each group and the variance or standard deviation (how widely distributed the scores are in each group) to see how much the two groups overlap. If they don't overlap very much, then the groups can be considered different from each other.

©iStock.com/MediaProduction

We have Guinness to thank for the statistic known as the *t*-test.

You might find it fun to know that we have beer (well, Irish dry stout to be exact) to thank for the invention of the *t*-test. William Sealy Gosset, a brewer at Dublin's Guinness Brewing Company, had to test the amount of stout in each batch of beer for quality control (Mankiewicz, 2004). It would have been impossible for him to sample from all of the thousands of casks produced every single day, so instead he took a random sample from the morning batches and a random sample from the afternoon batches and compared them to each other, to make sure they were the same. Gosset's invention of the math behind his comparison was published anonymously (under the fake name "Student") and is now used for much more than making sure our beer tastes great.

What if researchers want to compare three or more groups? For example, a study might be interested in analyzing relationships in each of the seven continents, to test for cultural differences. The principle for comparing multiple groups is the same as for comparing two groups.

For each group, we calculate the average score and the standard deviation, just like before, but when several groups are involved it's called an **analysis of variance**, or **ANOVA** for short. ANOVA tests will tell you whether at least one of the groups is different from the others, and additional follow-up analyses can tell you the details of which groups are different and how much they vary.

Patterns in a Single Group: Correlations

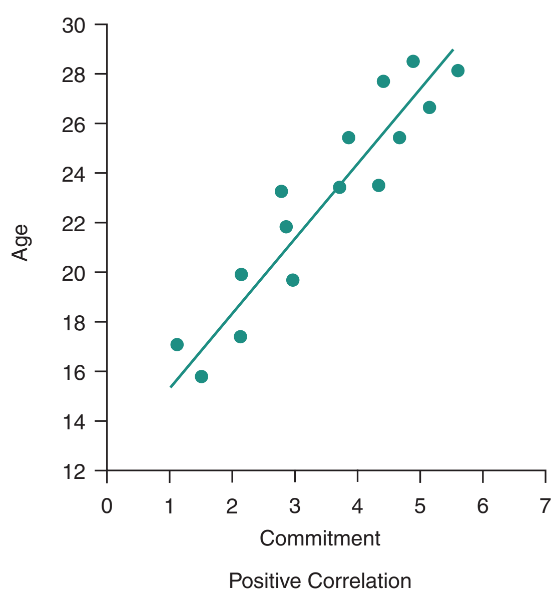
In Chapter 1, we talked about the difference between categorical and continuous variables. Variables that make different groups (Christians versus Muslims versus atheists, Canadians versus Brazilians versus Rwandans, people who listen to love songs versus a control group) are categorical and thus use *t*-tests and ANOVAs. The average scores and standard deviations of each group are compared to each other. However, many studies include a single group of participants and look for patterns of results among two or more continuous variables.

An example was covered in Chapter 1: the Alameda County Study (Berkman & Syme, 1979). There, social support was linked to mortality rates: People are less likely to die if they have a lot of social support. While *t*-tests and ANOVAs compare patterns of results in different groups, **correlations** look for patterns of results in a single group. Correlations test whether two different variables are systematically associated with each other, like social support and mortality were in Alameda County.

Correlations analyze the association between two continuous variables, meaning variables that have a range of scores that fall along a continuum. To test for a pattern, scores on each variable are gathered from as many people as possible and are then charted on a graph called a **scatterplot**. One variable is on the (horizontal) *x*-axis, and the other is on the (vertical) *y*-axis, and each dot on the scatterplot represents one person. Take a look at Figure 2.3 for an example (with theoretical data created for the purposes of this chapter—not from a real study).

The pattern shown in Figure 2.3 indicates that as people age through early adulthood, they are more likely to be in relationships with higher levels of commitment. The line summarizes the trend in the data. When a correlation is calculated, the number you get is called a **correlation coefficient**. It will always be a number between –1.00 and +1.00. How can you tell what the coefficient means? It's basically like a two-part code you can crack to understand what the pattern looks like on a scatterplot. There are two parts to the code: (1) the sign or direction (positive or negative) and (2) the number.

FIGURE 2.3 • An Example Scatterplot Graph



In this fictional study, age and commitment go together: As people get older, their commitment also increases. That means there is a positive correlation between age and commitment.

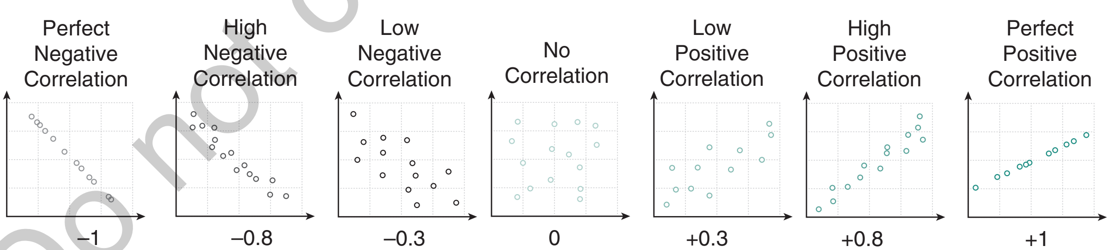
First, the sign will always be either a positive or a negative (unless the correlation is exactly zero). A **positive correlation** (between 0.00 and +1.00) indicates that both variables move in the same direction. In other words, if scores or values on one of the variables go up, values on the other variable will also go up. If one goes down, the other will go down. The example in Figure 2.3 shows a positive correlation: As age goes up, so does relationship commitment. Positive correlations are shown in scatterplots when the pattern or summary line moves from the bottom left-hand corner to the upper right-hand corner.

In contrast, a **negative correlation** (between 0.00 and –1.00) indicates that the variables move in opposite directions. As one variable goes up, the other goes down. For example, in Alameda County, more social support was associated with lower mortality (death rates). As social support went up, likelihood of death went down. Negative correlations will be shown in scatterplots with a pattern that goes from the upper left-hand corner to the bottom right-hand corner.

The second part of a correlation coefficient is the number, which will always be between zero and one (either positive or negative). The number tells you how clear the pattern is on the scatterplot, or how well the different dots (which represent people) fall along the summary line. Basically, this number tells you how much variability there is in the data, or whether some people don't fit the pattern. In Figure 2.3 you can see that not all of the people fall exactly on the line. If the dots all fall *exactly* on the line, meaning the pattern is perfect, the number you get will be 1.00. As the number gets closer to zero, it means the pattern becomes slightly less clear.

Note that coefficients of +1.00 and –1.00 are *equally strong*—both indicate perfect patterns, with all of the dots exactly on the line. It's just that in one case the variables move in the same direction (+1.00), and in one case they move in opposite directions (–1.00). Figure 2.4 is a summary of how to understand correlations, showing a range of patterns that move from perfect and positive, through no correlation at all, to perfect and negative. A zero correlation coefficient means that the two variables have no relation to each other at all, such as relationship commitment and someone's height

FIGURE 2.4 ● Summary of Types of Correlation



Correlations always range from –1.00 to +1.00. The sign (positive or negative) indicates whether the two variables move in the same direction or in opposite directions. The number (from 0.0 to 1.0) tells you how well each data point fits onto a general pattern. If a correlation is zero, it means there is no pattern or association between the two variables.

Source: Heinzen and Goodfriend, 2018.

or love of chocolate. These variables are not associated with each other at all, so the scatterplot would look like a bunch of random dots.

A final note, but a very important one, is that *correlation does not imply causation*. Being older doesn't *cause* someone to be more committed to relationships. It's possible that being older leads people to feel more prepared to make life plans, or being older leads to social expectations to marry and start a family. But without doing an experiment with random assignment to different conditions, causal inferences shouldn't be made.

An interesting example of this principle is, again, the Alameda County Study (Berkman & Syme, 1979). There, having more social support was negatively correlated with death rates. But does having more friends and family around lead to heathier behaviors? Or is it the other way around? Maybe being healthier leads to better quality relationships because people have more energy, participate in more activities, have more money to spend on luxurious gifts for their friends, and so on. A good understanding of the limitations within each research methodology and statistical analysis helps in knowing what conclusions should really be made within each study that's done.

Interpreting "Dyadic" Data

There's one more important point to consider regarding analyses of study results about relationships.

Most statistical tests work with the assumption that each participant's scores are independent from everyone else's in the study. In other words, the participants haven't influenced each other. However, if a study includes two or more people who are in a relationship with each other, that assumption goes right out the window. Clearly, the happiness, satisfaction, and so on of one person in a friendship or romantic couple is likely to be influenced by the other person. So if a study wants to include *both* people, the methods and statistics get much more complicated.

Methodologically, researchers will have to consider the pragmatics of holding study sessions when both people can be there. That automatically makes things complicated, with new considerations such as whether they'll need to find a babysitter, whether they both have the same work schedule, and more. If participants are being compensated with something like extra credit, what if one person is in the class offering compensation and the other isn't? How will the other person be compensated? And if the study is longitudinal, what happens to compensation if the couple breaks up halfway through the study, or one person wants to stop participating and the other wants to continue? Researchers will need to consider these additional aspects when couples or friends are supposed to participate together.

Statistically, the analyses will also become more complicated. Data will now have to be analyzed knowing that the results of one person are linked to another, using techniques called **dyadic analysis** (Kenny, Kashy, & Cook, 2006). Essentially each *couple* will be analyzed and then compared to the other couples, instead of comparing individual to individual—so slightly different formulas need to be used for these studies. Dyadic analysis is needed for parent to child, friend to friend, partner to partner, or any other method that includes people who influence each other. All of these considerations add another layer of challenge to people who want to make sure the science of intimate relationships is done well.

CHECK YOUR UNDERSTANDING

2.4 Dr. Y conducts a study in which they ask people over 50 years old how many sexual partners they have had in their lifetime. Dr. Y then compares the answers based on participants' socioeconomic status: lower class, middle class, or upper class. Which statistical analysis should Dr. Y do to understand their results?

- **a.** Analysis of variance
- **b.** *t*-test
- **c.** Scatterplot
- **d.** Correlation

2.5 Dr. Z asks half of their participants to run on a treadmill for 5 minutes, and the other half of the participants listen to calm, soothing music. All participants then rate how attracted they are to photographs. To analyze whether physiological arousal influenced perceived attraction in this study, what statistical analysis should Dr. Z do to understand their results?

- **a.** Analysis of variance
- **b.** *t*-test
- **c.** Scatterplot
- **d.** Correlation

2.6 Dr. X finds that the more introverted someone is, the more likely that person is to say they have a high self-esteem. What can be safely concluded from the results of this study?

- **a.** Introversion and self-esteem are positively correlated.
- **b.** Introversion and self-esteem are negatively correlated.
- **c.** Being introverted causes self-esteem to go up.
- **d.** Both a and c are correct.

APPLICATION ACTIVITY

Try to draw scatterplots that show the following results:

- A study of 10 people that resulted in a correlation of −1.00
- A study of 15 people that resulted in a correlation of exactly zero
- A study of 20 people that resulted in a correlation of +0.75

CRITICAL THINKING

- Consider correlation coefficients. A common mistake people make when they are first learning about correlations is that positive correlations are somehow "stronger" than negative correlations, even if the number is the same (e.g., +0.8 and −0.8). Why do people tend to make this mistake?
- "Correlation does not imply causation" is easy to say, maybe, but often hard for people to really follow as a rule. Try to identify three examples of real-life correlations in the world around you (e.g., "Calorie intake and body weight are positively correlated"). For each, identify whether you think there is a causal connection between the two variables where one causes the other, or whether both variables influence each other mutually, or whether there is a third variable that might be involved.

Answers to the Check Your Understanding Questions

2.4 a, **2.5** b, and **2.6** a.

What Is "Open Science"?

Ethics are always important.

We've already discussed some ethical considerations, such as avoiding deception in studies whenever possible, making sure we get informed consent before participants start in a research project, and so on. The ethics of science are even broader, though, when we start to think about how studies happen from start to finish. What if a researcher misrepresented their results, or they decided to form a hypotheses only after they had already done analyses? What if they refused to share their data with other people, who could confirm the findings?

Open science is a movement to make scientific research transparent, accessible, cooperative, reproducible, and honest. The aim of open science is to remove barriers for the creation of studies, sharing of data and results, and analysis of implications or conclusions. It's a way of saying to others, let's all do this together in an open, honest environment. One specific goal is to increase the number of studies focused on replication of previous work, so we can be confident in the conclusions we make and in the theories we teach in classes and textbooks (like this one!). Replication of results is the topic of this chapter's "Research Deep Dive" feature.

RESEARCH DEEP DIVE

Getting It Right: The Role of Replication in Relationship Science

As scientists we are obsessed with getting it right. This impulse is not entirely about us being right in the sense of seeing our predictions gain support (though that is nice), but more about getting the facts straight. Make no mistake, the stakes are high. Published research becomes the foundation for policies, college courses, textbooks, general audience books, and life decisions. In all cases, people put their faith in science to help improve their relationships.

Science is the gold standard for establishing facts because it requires that several key criteria be present for a finding to be considered authentic. For example, scientific information must be falsifiable, which means there has to be some way to refute it or collect evidence that contradicts the alleged fact. Despite every scientist's best efforts, no study is 100% perfect. Thus there is always a chance a study's conclusions are wrong. Perhaps the most straightforward way to see if a statement of fact is false is to retest it. If someone else's research finds that a training program helps people find quality partners more easily, we need to be able to test that ourselves. That is, we need to check our work and test our theories.

A key way for relationship science (and all other scientific fields) to accomplish and demonstrate falsifiability is through replication, where scientists redo a study to determine if they get similar results each time. In fact, replication comes in a variety of forms, with each type along the continuum contributing to our knowledge base. The first type is a **direct replication** in which researchers

(*Continued*)

(Continued)

attempt to re-run the original study, sticking as closely as possible to the measures, manipulations, and/or procedures that other researchers used in the previous study. Direct replications help establish that a given finding exists. In other words, if two different research teams can obtain the exact same finding, then it gives us more confidence that the finding is real. Technically speaking, direct replications help falsify the null hypothesis (that there is no association between the variables in the study).

At the other end of the replication continuum is a **conceptual replication**, in which researchers study the exact same variables and test a similar association, but intentionally use different measures, manipulations, or procedures. Conceptual replications help establish the extent to which previously established associations between variables apply to other contexts. For example, if we find that being in love leads to higher relationship quality, is that true of each type of love? How about for other measures of relationship quality?

To make replication even easier, scientists are increasingly using more open science practices. It is important to realize that simply because replication in science in general is not perfect, it does not mean that it is not trustworthy. Even diamonds have their flaws. It simply means that you have to read scientific findings (as well as every piece of information you encounter) with a critical eye.

For more, read Lebel, P. E., Berger, Z. D., Campbell, L., & Loving, J. T. (2017). Falsifiability is not optional. *Journal of Personality and Social Psychology, 2,* 254–261.

There are several ways that open science encourages this kind of communication and exchange; a few are preregistration, results-blind peer review, and publication badges. We'll cover each idea below, but to learn more about this exciting trend in science, you can also go to the following websites:

- The Center for Open Science (https://cos.io/)
- The Open Science Framework (https://osf.io/)
- OpenScience (https://openscience.com/)
- ORION Open Science (https://www.orion-openscience.eu/)
- The FOSTER Portal (https://www.fosteropenscience.eu/)

Preregistration

Imagine that a scientist does a study in which they're not really sure what they're looking for or what the outcomes are expected to be. This is called exploratory research, and there's nothing wrong with it. But now imagine that after the results are analyzed, the scientist publishes the study and more or less pretends that they predicted the outcomes from the beginning. They look super smart! But it's not an honest approach.

Open science's solution is **preregistration**, a practice of specifying—in advance—your hypotheses, procedure, and statistical plan for analyses (see Nosek, Ebersole, DeHaven, & Mellor, 2017). This plan is made publicly available to anyone, so you are committing to everything in an open, transparent way. Several preregistration templates have been created to help people through this process, where researchers can post their plans on independent websites.

Preregistration is not without problems. For example, you might say that you're going to get 100 people for your study, but you can only get 75. Or you might assume that people will pay attention to instructions during your procedure, but some of them don't and they mess up what they are supposed to be doing. Or you might realize after you've collected data that you had a typo on your scale that changed what the question was asking. Scientists are certainly not perfect, and mistakes can be made. But all of these changes can simply be documented and explained. That way, readers of the research can understand exactly the process that occurred and why changes had to be made.

Typical questions you'll answer on a preregistration form are things like this:

- What are your hypotheses? If you're doing a quasi-experiment or experiment, what are the independent and dependent variables? If you're doing a correlation study, do you expect a positive or negative correlation?
- What exactly will the procedure be—what will participants do? What will be the order of procedural steps? How long will it take each person to do the study? How will you do random assignment (if relevant)?
- How will you recruit participants, and how many do you expect to find? Will anyone be excluded from data analysis—and if so, why?
- How will each of your variables be operationalized and calculated (if you're doing a quantitative study)? What statistical tests will you use to analyze the results?

Results-Blind Peer Review

Every academic field has professional journals, where researchers publish their results.

Most of these journals are what we call "peer-review" journals. That means that before any article is accepted for publication, it's sent out to other experts on that topic to see what they think. Those people, called reviewers, give the author(s) anonymous feedback about whether they think the article is worthy of being published. Sometimes reviewers will make suggested changes that they want to see; if those changes are made, the journal will usually publish the article. Sometimes, however, the reviewers can simply say that they don't like the study and stop it from being published.

Until the open science movement, all of this reviewing happened after a study was completed and written up. That meant that the peer reviewers knew how the study

FIGURE 2.5 • The Results-Blind Peer Review Process

When an article goes through the "results-blind peer review" process, outside experts give feedback about the quality and importance of an article before the data are actually collected. Then, they review a second time, focusing on whether the study followed the original design plan.

Source: Center for Open Science, used under Creative Commons Attribution 4.0 International license, https://creativecommons.org/licenses/by/4.0/.

turned out. The problem with this is that it can lead to biases in what is and isn't accepted for publication. Maybe the reviewers wouldn't like the results because they go against a theory they favor. A more common problem is that studies usually weren't published if their results didn't show statistically significant findings or results that matched their hypotheses.

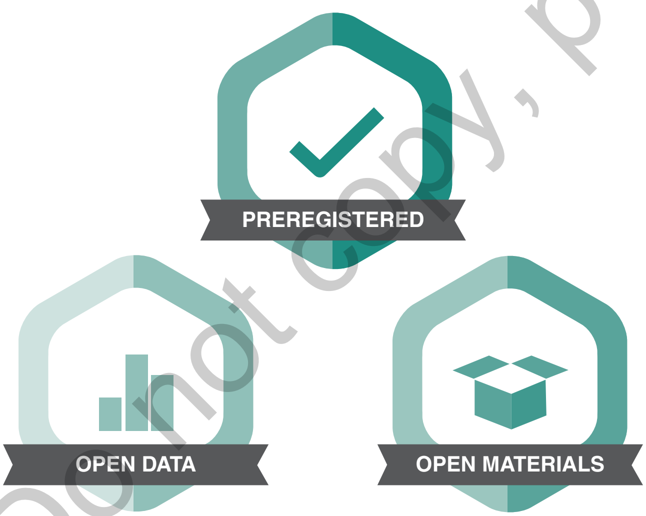
These problems can largely be eliminated with a practice called **results-blind peer review**, which means that reviewers are asked about the importance of the study *before* they see the statistical outcomes, as shown in Figure 2.5. If they agree that the study has merit, they accept it for publication at this stage. Reviewers will also be asked for their feedback after the results are calculated—but now they comment on whether the study followed the preregistration plan and interpreted everything correctly. That way, even if the results surprise everyone, the study still gets published. Chris Chambers, the chair of a committee at the Center for Open Science, stated the benefits of this process like this: "The incentives for authors change from producing the most beautiful story to producing the most accurate one" (Center for Open Science, 2020b). Just like a relationship partner, science is even more beautiful when it's accurate and honest.

Publication Badges

Beyond the rewards of knowing you're doing good science, what incentives are there for people to engage in open science practices?

One reward is the use of **badges**, or visual icons that mark a study with signals that it has followed these procedures. You can see what the badges look like, at least for some journals, in Figure 2.6. If a study followed the requirements for each or all of the badges, the badges it earned will appear on the first page of the published article. For example, if they posted their original, raw data spreadsheets online, they get the "Open Materials" badge. Over 50 journals now use the badge system, and early trends show that they really do increase the number of scientists who share their data publicly (Kidwell et al., 2016; Rowhani-Farid, Allen, & Barnett, 2017). The open science movement is likely going to increase in usage and popularity over the next several years, as many people see it as the only way to make the scientific process truly objective and transparent.

FIGURE 2.6 ● Examples of Open Science Badges



Professional journals are increasingly marking studies with these images, called "badges," when they follow open science guidelines. These examples are from the Center for Open Science.

Source: Center for Open Science, used under Creative Commons Attribution 4.0 International license, https://creativecommons.org/licenses/by/4.0/.

CHECK YOUR UNDERSTANDING

2.7 The movement to make research more transparent, accessible, cooperative, and honest is called:

a. Academic authenticity
b. Public futurism
c. Open science
d. Methodological truth

2.8 Dr. X wants to replicate the idea behind someone else's study but to use new procedures and a different sample of people to see if the phenomenon holds up again under these new methods. What kind of replication would Dr. X's study be called?

a. Abstract replication
b. Conceptual replication
c. Multiple replication
d. Direct replication

2.9 Research on the use of "badges" by journals has shown that badges:

a. Increase the number of scientists interested in publishing their work
b. Decrease external validity in most studies
c. Increase the number of scientists who publicly share their data
d. Decrease the amount of deception used on participants in studies

APPLICATION ACTIVITY

While many people praise open science practices, they are not without drawbacks. One article that discusses some disadvantages can be found online by searching for the title "Open science isn't always open to all scientists" (Bahlai et al., 2019). Read this article and discuss or write your own opinion about the pros and cons of the open science movement.

CRITICAL THINKING

- Some professional journals charge for copies of their articles, or they require people to pay for subscriptions. Others offer their articles to readers for free, but they require that the scientists themselves pay to publish their work in the journal. What do you think is the best system for research to be available to other scientists or the general public, in terms of how it is funded? Should there be a new system, like a "science tax" that everyone pays but is used to make scientific progress available to everyone? Discuss how you think science should be funded, and why.
- The peer review process, even when done using the "results-blind" procedure, can be frustrating to people who want to publish their work. Sometimes, it seems unfair that anonymous people get to judge your work and decide whether it's "worthy." On the other hand, peer-reviewed articles are considered more credible because they have passed this hoop of acceptance. Do you think the peer-review system is good or bad, and why? If you don't like it, is there a better alternative?
- If you were a researcher, would badges make you more likely to engage in open science practices? Why or why not?

Answers to the Check Your Understanding Questions

2.7 c, **2.8** b, and **2.9** c.

Chapter Summary

Big Questions
1. What research methods are used to study relationships?
2. How are results analyzed?
3. What is "open science"?

What research methods are used to study relationships?

The scientific method consists of steps: observe a pattern, generate a hypothesis, test the hypothesis, and interpret the results. It then continues in a circular, repeatable cycle. Good research has high internal validity, external validity, and replicability. Researchers choose between a qualitative study, which gathers open-ended data through interviews or surveys, versus a quantitative study, which gathers data in numerical form (e.g., a scale that ranges from 1 to 7). In addition, several methods are used. These included archival research (or information originally gathered for another purpose), naturalistic observation (watching people in their natural environments), self-report surveys, quasi-experiments, and experiments. Each method has advantages and disadvantages. Finally, ethical treatment of participants includes practices such as obtaining informed consent, avoiding deception when possible, giving them the right to withdraw, and performing a thorough debriefing.

How are results analyzed?

When two groups are going to be compared to see if they are different from each other, the statistic used to analyze data is called a *t*-test. It compares the two groups' average scores and standard deviations, to see how much they overlap. The same principle is used to compare three or more groups, but then the statistical test is called an analysis of variance, or ANOVA for short. When a single sample is used in a study, but researchers want to test for associations between variables, correlation tests are used. Correlations can be positive (meaning both variables move in the same direction) or negative (the two variables move in opposite directions). The number for a correlation will always be between zero and one, with higher numbers meaning a stronger association between the two variables. Importantly, just because two variables are correlated with each other, it doesn't necessarily mean that movement of one causes movement in the other.

What is "open science"?

"Open science" is a movement to make research more transparent, cooperative, and honest. It involves practices like preregistration, where researchers specify their hypotheses, procedures, and statistical plan for analyses before any data are actually gathered. Another practice in open science is results-blind peer review, where other experts judge the value and quality of a study without knowing what the results were, so they can't be biased by the outcome of the study. Finally, many journals are now awarding "badges" for people who use open science. Badges are icons that appear on the first page of a published study that indicate the usage of various open science practices within a given study.

List of Terms

Learning Objectives	Key Terms
2.1 Describe the scientific method, different types of research study, and ethical considerations.	Scientific method Hypothesis Internal validity External validity Replication Qualitative study Quantitative study Cross-sectional study Representative sample Random sampling Archival data Naturalistic observation Reactivity Participant observation Self-report survey Social desirability Quasi-experiment Experiment Random assignment Experimental group Control group Independent variable Dependent variable Institutional review board (IRB) Informed consent Deception Debriefing
2.2 Compare and contrast the most common ways to analyze and interpret research results.	*t*-test Analysis of variance (ANOVA) Correlation Scatterplot Correlation coefficient Positive correlation Negative correlation Dyadic analysis
2.3 Analyze the "open science" movement and what it means for future research endeavors.	Open science Direct replication Conceptual replication Preregistration Results-blind peer review Badges