STATISTICS: OPEN FOR EVERYONE

Christina R. Peter
Los Angeles Valley College

LibreTexts™

Statistics: Open for Everyone

Christina R. Peter

Los Angeles Valley College

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

This text was compiled on 03/18/2025

TABLE OF CONTENTS

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing****.*

CHAPTER OVERVIEW

1: Introduction to Statistics

1.1: What is Statistics?

Statistics is a field build from the foundations of mathematics as a tool to answer questions about nearly any aspect of reality. The term "**statistics**" generally refers to all the field-specific procedures and formulas that can be used to summarize information gathered from samples or to test hypotheses and assumptions with sample data. The term "**statistic**" is used to refer to the summary or outcome of those procedures. Therefore, the term "statistics" is also used as the plural form of "statistic" to refer to multiple summaries of outcomes. Thus, in statistics (the field), we use statistics (meaning the procedures from the field of statistics) to yield statistics (summaries of outcomes of those procedures).

Statistics includes many procedures which generally serve one overarching goal: to summarize or understand what is likely true based on incomplete information. To get there, statisticians start by using mathematical techniques to first summarize the data they have. This form of statistics is referred to as **descriptive statistics**. Then, when desired and possible, more advanced techniques are applied to test what is likely true beyond those data based upon those data. This form of statistics is referred to as **inferential statistics**. Most of us have taken math classes in the past where we learned about things like adding, subtracting, multiply, and dividing. These are all part of a type of math called arithmetic. One great thing (or frustrating thing, depending on how you view it) about this kind of math is that there is often one correct, knowable answer. For example, we know that 3 + 2 = 5. The answer is knowable. We even get to learn about "proofs" in some more advanced math classes. When we say something is proven we mean it is known and true.

However, there are times when we have incomplete information and the pieces of information we do have don't all tell the same story but we still want to use them to understand what is generally or likely true about the world. Statistics was developed to address these situations. Statisticians still use those mathematical foundations and operations (such as adding and dividing) but develop and implement specific techniques and language to allow them to summarize data and estimate what is likely true rather than what is absolutely known to be true. To understand the kinds of situations for which statistics were built and how the field functions, we can start by reviewing some key tenets of statistics.

> *Statistics is the study of data from samples.*

1.2: Tenets of Statistics

There are three main tenets of statistics: Variability, Probability, and Uncertainty. Let's review some key terms and concepts that connect to and help us understand these three tenets.

Variability

In statistics, data are collected about variables. **Data** is the plural term used to refer to multiple bits of information together. The term **data set** can also be used to refer to data from one or many variables. Datum is the singular form of data which is used to refer to one piece of information. **Variability** is a broad term that refers to the fact and ways in which things can differ. A **variable** is anything that is measured which is not always the same. For example, we could collect data about the rainfall in inches for each day of January. The variable measured would be rainfall. Though it is possible for the amount of rain measured each day to be exactly the same, it can (and likely will) be at least a little different on some days compared to others. Thus, the thing being measured (rainfall) is considered a variable. The data for this variable would be all the numbers for the amounts of rainfall each day. For example, it might have rained 0.00 inches on the 1st, 1.25 inches on the 2nd, 0.40 inches on the 3rd , 1.25 inches on the 4th and so on. These numbers (0.00, 1.25, 0.40, and 1.25, respectively, are the first four pieces of data in the data set).

In contrast, a **constant** is anything which is measured that cannot or does not vary. Whenever something being measured is always the same or did not differ at all among the instances being measured, it is considered a constant. For example, if a person grew up in a town called Statistonia, and each day you collected data on where they grew up, the data would be the same every day. Every day the data for the variable *Childhood Town* would be *Statistonia*. It doesn't matter how many times you ask it, the data are always the same so there isn't much you can do or say with these data. This makes constants quite limited in their utility and, thus, they are rarely of interest to statisticians when testing hypotheses.

Instead, statisticians are particularly interested in examining data collected about variables. The nature of variables means that they might have interesting patterns on their own or in relation to other variables. Let's consider an example. We can measure the variables income and happiness to see if there is any pattern between them. If we measure income and level of happiness among many people, we will likely see that income differs from person to person and that happiness also differs from person to person. Some incomes may be quite low relative to others while others are quite high. We might see a pattern where incomes are generally lower than $60,000 but that a few people make well over $100,000. We would likely also see variability in happiness. For example, if individuals rated their happiness on a scale from 1 (not at all happy) to 10 (extremely happy), we might see that most people rated their happiness somewhere between a 4 and a 6 but that a few people rated their happiness below this range and a few others rated their happiness above this range. With these kinds of descriptions, we are summarizing some patterns we see in the data for each variable. Already, this is interesting but let's take it a step further. We may be curious whether income relates to happiness such as whether happiness tends to be higher when income is higher. Because both things which were measured varied, statistical techniques to assess this proposed relationship are possible (and we will learn about them and how to use them in Chapter 12).

However, patterns like these cannot be discerned when looking at constants on their own nor in relation to variables. Let's take the same example but consider what would happen if income was a constant. Suppose that when the data were collected, income was a constant $60,000 meaning every person made this amount. There isn't much we can say other than "the income was a constant $60,000." That's it. Even if happiness varied, we couldn't say whether there was a relationship between income and happiness because such a pattern cannot be established or tested when either or both of the things which were measured were constants; we can only establish or test patterns between and among variables. Therefore, the field of statistics is primarily focused on variables.

Probability

In statistics, data are collected from samples. Thus, we can say that statistics is the study of data from samples. However, these data are often collected from samples in the hopes of understanding populations. The distinction between populations and samples is, therefore, important for understanding why and how probability is central to much of statistics.

Populations

A **population** includes all cases that comprise a specified group. The term **case** is used to refer to a single member of a population. Another way of saying this is that a population includes all the examples of who or what we are trying to understand. For example, a statistician might be interested in understanding college students. The defining characteristic that identifies who the statistician wants to understand is "college student." Therefore, the population would be comprised of every college student that exists. However, it is often difficult and is sometimes impossible to collect data from a population. Think of how difficult it would be to identify and collect data from every single college student that exists. Even if we could identify and locate every college student, for practical and ethical reasons, we are unlikely to be able to get data from all of them. If even one college student cannot be located or doesn't want to provide information about themselves, the data would be incomplete. Therefore, it is far more common to have data from samples than from populations.

Samples

A **sample** includes some, but not all, cases that comprise a specified group. Another way of saying this is that a sample includes some of the cases from the population we are trying to understand. By definition, then, sample data are incomplete because data are not available and known for at least one member of the population of interest. Depending on the size of the population and the ability to identify each case and collect data, data may only be available from a relatively small sample of population members. For example, though there may be millions of college students, data might only be collected or available from 250,000, 10,000 or maybe even only 50 of them. Each of these would be samples of the population of college students. When a result is summarized from the full population, it is referred to as a **parameter**, similar to how a result summarized from a sample is referred to as a statistic. You can remember these terms alliteratively: Populations provide parameters while samples submit statistics. Because statistics are yielded from sample data, which are incomplete representations of population data, statisticians can only deduce what is probably true about populations rather than what is definitely true. Therefore, the theme of probability appears throughout statistics.

Probability refers to how likely something is to occur or be true. Because data from samples are both more available than those of populations and have the limitation of being incomplete, the field of statistics developed techniques that use probabilities to estimate what is true based on sample data. These techniques and, thus, statisticians, will always have to deal with some ambiguity. Statistical procedures are used to estimate the probabilities that various observations in sample data represent realities in the populations. Therefore, probability is central to statistics.

Some statistical procedures are built from assumptions about the patterns of data in populations. These are referred to as parametric statistics. **Parametric statistics** refer to techniques that use data from samples that are assumed to have been drawn from populations which are distributed in specific ways. These distributions are known as normal distributions which we will learn about in detail in Chapter 5. Essentially, this means that the data are assumed to follow certain patterns in the population and that probabilities can be calculated based on those assumptions. We will learn how to use some parametric statistics in later chapters (such as Chapters 8, 9, and 10). There are also non-parametric statistics. **Nonparametric statistics** are analytic techniques that are used when data are not assumed to follow the normal distribution a priori. Instead, these techniques are used when either the assumptions of parametric statistics are violated or cannot hold true because of the types of data being used. We will learn about one such test in Chapter 14. For now, it is important to know that probabilities are important components of both parametric and non-parametric statistics.

> *There are no guarantees in Statistics.*

Uncertainty

Because statistics yields estimates of probabilities, it means there will always be some uncertainty. The fact that statistics is focused on data which vary and are incomplete (because they come from samples rather than populations), means that statisticians focus on what is probably true and cannot deduce what is definitely true. Statisticians, instead, use procedures to deduce what is likely true about populations and their parameters but cannot 100% guarantee that these things are true. For these reasons you will notice that statisticians often use hedging language. For example, a statistician may have strong evidence suggesting that those who study more tend to earn higher grades. However, even if the data are quite strong, they cannot guarantee that all individuals who study more will earn higher grades. Therefore, they are likely to say something like "Studying more tends to produce higher grades" or "Increasing your hours of studying will likely improve your grades" rather than something absolute such as "If you study more, you will get higher grades." This form of hedging language is an important way of communicating that every analysis has some limitations, that nothing in statistics is a guarantee, and that those who are skilled in statistics are always open to the possibility that there are disconfirming data out there.

Uncertainty is an important reality of working with data and statistics that some people find frustrating, and for good reason. Of course, we would ideally like to have certainty and guarantees over probability but that is not always possible. This does not mean, however, that the field has nothing to offer and that we should discard our book and abandon our study of statistics here. Here's why: knowing what is probably true has value and can be useful in guiding decision-making.

Let’s take a quick but dramatic example. Suppose someone has a serious illness where the chance of dying within 6 months is 80% if left untreated. The 80% is a descriptive statistic that summarizes a probability. Suppose that two medications are available to treat the illness and a doctor must decide what to prescribe. Suppose the risk of death for those who took Medication A in a large study was only 30% and the risk of death was 80% in those who took Medication B, with no side effects reported in either group. Though there is still uncertainty because nothing in this example states what occurs 100% of the time, there is still information here that is useful for the doctor and patient as they make decisions about treatment options. Additional information about how similar the patient is to those in the study and statistical findings for other potentially relevant factors and studies would also be used to determine the treatment plan that is most likely to benefit the patient. When certainty is not possible, probability is the next most useful option and is more beneficial than being tasked with making decisions in the absence of any information.

Reading Review 1.1

1. What is the overarching goal of statistics?
2. Which tenet refers to the fact that statistics focuses heavily on data collected about things which are not the same for all cases?
3. Statisticians focus on data from ______, which include some, but not all, of the cases of interest.

1.3: Types of Variables

All things about which data are collected are variables, however, not all variables are measured or represent the world in the same way. There are several ways variables can be defined and categorized based on the nature of the variable itself and its operationalization. To **operationalize** is to define a variable for the purposes of measurement. Therefore, when a researcher or statistician operationalizes, they are defining the way a variable will be measured to gather data for their research. The nature of variables limits how they can be operationalized and measured. The way they are measured then limits how they can be analyzed. Therefore, it is important to understand the different kinds of variables and ways they can be categorized.

Qualitative verses Quantitative

The first main way to categorize variables is by whether they are qualitative or quantitative. **Qualitative variables** are those which vary in characteristic, category, type, or kind rather than amount. Eye color is qualitative because we use categories to define the type of color each individual's eyes are. Major in college is also qualitative because names are used to differentiate among the majors. When someone says their major is Psychology and someone else says their major is Nursing, the two individuals vary in type of major rather than amount of major. We wouldn't say that one of this is more or less "majory" than the other because the names of the majors indicate differences in type of major not amount of major. Thus, eye color and major are two examples of qualitative variables.

In contrast, **quantitative variables** are those which exist and are measured in amount. Weight is a quantitative variable because values are used to represent differing amounts of weight. When a variable is quantitative, comparisons of amount are possible. For example, we can say that a cat that weighs 19 pounds is heavier than a cat that weighs 14 pounds. Here the word "heavier" indicates a comparison of the amount of weight between the two cats. Many other variables are quantitative such as sleep measured in hours, distance driven measured in miles, or math knowledge on a 10 point quiz. In each of these examples you can see what the variable is (sleep, distance driven, math knowledge) and how each is being quantitatively operationalized (as hours, miles, or points, respectively).

> *When something varies in characteristic, category, type, or kind, it varies qualitatively.*
> *When something varies in amount, it varies quantitatively.*

Discrete verses Continuous

There are a variety of ways quantitative variables can exist and, thus, be properly measured. One way they can differ is by being discrete variables or continuous variables.

Discrete variables are sometimes referred to as discontinuous variables. When a quantitative variable is **discrete** it means it exists and can be measured in counts. The data for discrete variables are often whole numbers. For example, the number of children someone has is a discrete version of a quantitative variable. Number of children is quantitative because data are being collected to represent an amount. This amount is discrete because children can only exist in whole units and are not divisible into parts of a whole. You can have 0 children, 1 child, or perhaps 12 children but you cannot have 1.50 children. Discrete variables can be thought of as thing that can only vary in specific counts rather than any amount.

In contrast, when a quantitative variable is **continuous**, it means it can exist and, thus, be measured in infinitely differentiated amounts. Hours of sleep is continuous because the amount of time someone sleeps can be exactly 7.00 hours or 8.00 hours, for example, or it could be anything between those such as 7.50 hours, 7.25 hours, or even 7.9999 hours if we have a very accurate and specific timer. When data are continuous it means that it is possible that amounts vary at or between integers.

Scales of Measurement

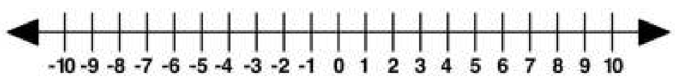
Another way to distinguish among types of variables and how they are measured is through the scales of measurement. When a variable is operationalized, one of four scales of measurement can be applied. The four scales of measurement are: ratio, interval, ordinal, and nominal. These are the categories for the four different ways things can be measured. The first two scales of measurement (ratio and interval) differentiate between the two common ways something quantitative can be measured. The fourth scale of measurement (nominal) is used when something qualitative is measured. The third scale of measurement, the ordinal scale, is best understood as existing in the tension between what is quantitative and what is qualitative.

The Ratio Scale

The ratio scale is the most specific and least limited of the quantitative scales of measurement. The **ratio scale** organizes amounts that increase by the same magnitude throughout a sequence with a true zero used to represent a complete absence of that which is being measured. Thus, ratios scales have two defining features: 1. A true zero and 2. Equally sized changes in amount between each whole number. Think of the ratio scale as referring to anytime something is measured based on where it would fall on the right side of the number line, starting from 0. On the number line 0 represents a complete absence of something. As the amount of the thing being measured increases, values move to the right on the number line. This scale of measurement is the one that is often presumed when considering real world values in math. When we are given a problem such as "If you have three apples and get two more apples, how many apples do you have?" the ratio scale is presumed used.

The word "ratio" is given to these scales to indicate that they have a feature no other scale of measurement has: the ability to make meaningful comparisons with ratios. A ratio refers to the relative size between two amounts. Another way of saying this is that a ratio tells how many times one amount is a multiple of the other. These kinds of relative comparisons can only be made when a true zero exists and is known based on the measure used. Let's consider ratios using coffee measured in ounces. The least amount of coffee one could have is 0 ounces. Someone who has 16 ounces of coffee has twice as much as someone who has 8 ounces of coffee. We can make this statement because coffee is being measured on a ratio scale. Hours of sleep also uses the ratio scale and, thus, comparisons of relative amount can be made. We can correctly state that 4 hours of sleep is twice as much as 2 hours of sleep or that 5 hours of sleep is half as much as 10 hours of sleep.

This may seem like an absurdly obvious thing to say because, of course, 4 is twice as much as 2 and 5 is half as much as 10. The reason this seems so obvious is because most of the mathematical concepts we know about the world are focused on things which exist on the ratio scale of measurement. We rarely overtly learn or think about quantitative things which are not on this scale of measurement. We are accustomed to thinking of 0 as representing a complete absence because that is its mathematical role. That is what it is meant to represent. Therefore, we often treat positive numbers as though they represent the relationship of those numbers to 0 on the number line. We may not have ever considered "what if there was no 0" or "what if 0 wasn't really known to be 0." However, sometimes 0 is not truly knowable based on the way something can be measured. In those cases, different scales of measurement are used and ratio-based statements of comparison are not appropriate.



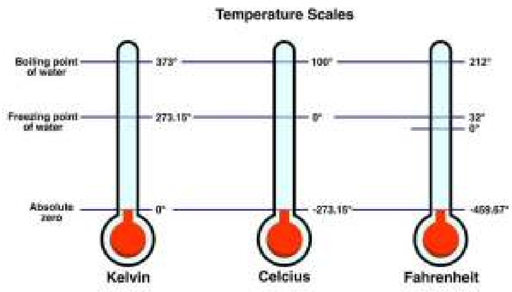
The Interval Scale

The interval scale is a quantitative scale of measurement without a true, known zero. **Interval scales** organize amounts that increase by the same magnitude throughout a sequence but do not have a true, known zero point. Thus, they are quantitative but only meet one criteria that is also met by the ratio scale: Equally sized changes in amount between each integer. Integers refer to whole numbers (such as 3) or their negatives (such as -3). An interval is an equally sized segments of change in magnitude. This scale of measurement was named for this defining feature.

Think of the interval scale as being a segment of the number line without starting at the true zero. Equally sized differences between numbers on the scale reflect equal differences in magnitude. However, the true zero is either not known or is not represented as "0." Many interval scales do not use 0 because the value chosen for 0 would be somewhat arbitrary. When 0 is used on an interval scale, the statistician must keep in mind that this value does not indicate a known and complete absence of that which is being measured. Some interval scales have a zero but also have negative numbers. Ratio scales cannot measure into the negative because you cannot physically have less than nothing (or absolute zero) of something.

Comparing the Ratio and Interval Scales

When an interval scale is used, statements of ratio are not appropriate. Let's explore why with an example. A good example of something that is measured on an interval scale is temperature in Fahrenheit. Temperature is quantitative and exists in a continuous way. However, there are lots of ways it can be operationalized. We could, for example, measure temperature using Fahrenheit, Celsius, or Kelvin. These are three different quantitative temperature scales. Fahrenheit and Celsius are interval scales of measurement. Each has a 0 but it is not a true zero because it does not represent a complete absence of heat. If it did, then it would not be possible to observe of experience a negative temperature on either scale.



When each of these two interval scales were created, a relevant point of temperature was labeled as 0°, but neither was an absolute zero. Anders Celsius developed the Celsius scale and chose to label the temperature at which water froze as 0°. Daniel Gabriel Fahrenheit developed the first version of the Fahrenheit scale and chose to label the temperature at which a saltwater mixture froze as 0°. Thus, each of these scales call different amounts of heat 0 and it is possible to experience temperatures below 0 degrees on each of these scales. This is because the point that was chosen to be labeled as 0 on each of these scales does not represent a complete absence of heat. Because of this, it is not accurate to state that 80° Celsius is twice as hot as 40° Celsius nor is it accurate to state that 80° Fahrenheit is twice as hot as 40° Fahrenheit. The only way these kinds of ratio statements can be made is if the 0 on the scale represents a true zero.

However, temperature can be measured with the ratio scale known as Kelvin. Kelvin is a way of measuring temperature that counts up from a point known as absolute zero. There are no negative values on the Kelvin scale. Absolute zero, which is 0 degrees Kelvin, represents the absence of heat; it is the lowest possible point determined for heat and the Kelvin scale counts up from this point. Because of this, the Kelvin scale meets both of the criteria to be a ratio scale of measurement.

When data are collected on a ratio scale, it is possible to make ratios. For example, 80° degrees Kelvin is twice as hot as 40° degrees Kelvin, something we could not say for these temperatures on the Celsius or Fahrenheit scales. This is because comparisons of ratio can only be made when a quantitative scale of measurement with a true zero, like the Kelvin scale, is used. Fortunately, however, there are few other limitations to using interval scales and, thus, data which were collected using interval or ratio scales of measurement can often be analyzed using the same techniques.

The Ordinal Scale

The **ordinal scale** is used when something can be organized into quantitative sequences without specifying the precise magnitude of differences in the sequence. Thus, when things can be put into ranks or orders of magnitude, they are on the ordinal scale. For example, the speed of runners can be ranked from fastest to slowest or from first place (representing fastest) to last place (representing slowest). Ordinal scales are limited in that the magnitude of change between ranks is not necessarily, or known to be, consistent. When numbers are used to specify who is the fastest (1st) to who is the slowest (5th) among five runners, for example, the numbers represent the five ranked or ordered places. However, the difference in how much faster first place was from second place is not necessarily the same as how much faster second place was from third, third was from fourth, or fourth was from fifth. Height can also be measured on the ordinal scale by organizing persons from shortest to tallest. In this way, the heights of five people could be ranked from 5 (tallest) down to 1 (shortest). However, the amount that each person's height differed from the next tallest person in the ordered sequence is not necessarily the same because the specific amount of change in height from person to person is rarely stable. The tallest person may be 1 inch taller than the second tallest person but the second tallest person might be 3 inches taller than the third tallest person.

The ordinal scale is unique in a few ways. First, it is the only scale of measurement where words or symbols are often used to represent quantity instead of numbers. Words like “first,” “tallest,” “most,” “last,” “shortest,” and “least” are used for some variables measured using ordinal scales. Second, it is the most limited of the three quantitative scales of measurement because it is the least specific in its ability to differentiate amounts. This limits the way the data can be computed and the kinds of quantitative comparisons that can be made. In fact, ordinal data should generally be analyzed using techniques appropriate for qualitative data rather than those that can be used to quantitative data. For this reason, the ordinal scale is sometimes argued to be a qualitative scale of measurement despite it being quantitative in nature (because it organizes data based on magnitudes). Fortunately, many quantitative variables that are sometimes measured on the ordinal scale, such as speed or height, can be measured with more specificity using the ratio or interval scales as long as the researcher or statistician operationalizes them that way. When using a ratio or interval scale is possible, it is often best to use that instead of an ordinal measurement.

The Nominal Scale

The nominal scale of measurement is the category used for all qualitative variables. The nominal scale is used when observations are labeled and categorized to differentiate them by characteristic. “Nom” means name and “nominal” roughly means “in name” or differentiated in name or type. Qualitative variables such as eye color and major, for example, are measured on the nominal scale. Data for variables measured using a nominal scale can be referred to as nominal data.

Connecting Topics

Interval and ratio data are often a good fit for parametric statistics while nominal and ordinal data often require non-parametric statistics.

Nominal data can be represented using words or names, symbols, or numbers. However, it is important to keep in mind that when numbers are used for nominal data, they function as individual or category names rather than indicators of quantity. An example of this is an ID number. An ID number is a set of digits used as a name for an individual or item but does not represent an amount. If one person has a student ID number of 80645 and another has a student ID number of 29448 it does not mean that one of them is more of a student than the other. The numbers are simply used the same way as names to differentiate between the individuals.

Independent Variables verses Dependent Variables

So far we have been considering variables one at a time. When one variable is considered or summarized at a time, we refer to the procedures and results as **univariate statistics**. “Uni” refers to one of something and “variate” means the something being referred to is a variable. However, it is also possible to consider two variables at a time and how they might be related or connected. When the associations or connections between two variables are considered, the procedures and results are referred to as **bivariate statistics**. “Bi” refers to the co-existence or consideration of two things at once. It is also possible to focus on two or more variables at which time the procedures and results can be referred to simply as **multivariate statistics**. For the first several chapters of this book, we will focus on univariate statistics (such as Chapters 2, 3, and 4). However, in later chapters we will also explore bivariate statistics, some of which (such as Chapters 8 and 9) further define variables as independent or dependent.

Statisticians sometimes use bivariate analyses to see if one variable is the cause of change or differences in another variable. When this is done, the variables are differentiated using the terms the independent variable and the dependent variable. An **independent variable** is the assumed or hypothesized cause in a cause-effect relationship. A **dependent variable** is the assumed or hypothesized thing which is affected in a cause-effect relationship. Whether something is considered an independent variable or a dependent variable is specific to what is being considered or tested at that moment. This means that it is possible for the same variable to sometimes be an independent variable and other times to be a dependent variable. Take grades as an example. If we want to test whether studying causes grades to be higher, studying is the independent variable and grade is the dependent variable. However, if we want to test whether grades cause students to be considered more desirable job candidates, grades are now the independent variable and desirability of the candidate is now the dependent variable.

It is important to note that not all variables are either independent or dependent. These designations are only used when cause-effect connections between or among variables are being considered to tested. In the absence of those scenarios, these terms should not be applied. For example, a statistician may want to test whether hours of sleep relate to amount of caffeine consumed where the statistician is either not interested in, cannot deduce, or is not considering, whether one variable is the potential cause of the other. In these situations, the terms independent variable and dependent variable are either not used or are used with the understanding that they are being theoretically positioned as cause and effect thought they cannot be fully tested as such. We will see some examples of these scenarios in later chapters (such as Chapter 12).

Reading Review 1.2

Use each description of a variable and how it was measured to categorize it. Variable names are underlined to help you identify them.

1. Ethnicity measured by having participants fill in the ethnicity with which they identify.
   a. Is the variable quantitative or qualitative?
   b. Which scale of measurement is being used?
2. Chips consumed measured in ounces.
   a. Is the variable quantitative or qualitative?
   b. Which scale of measurement is being used?
3. Shirt size measured in sizes of XS, S, M, L, XL, and XXL.
   a. Is the variable quantitative or qualitative?
   b. Which scale of measurement is being used?

Identify the independent variable and the dependent variable in each bivariate sentence below:

4. Exercise decreases stress.
   a. Independent variable:
   b. Dependent variable:
5. Popularity is impacted by behavior
   a. Independent variable:
   b. Dependent variable:

1.4: Types of Statistics

Statistical procedures were each developed to address specific kinds of data, questions, and/or hypotheses. Because there are so many things researchers and statisticians want to know about the world, there are many different techniques, procedures, and formulas that have been developed. Despite the complexity and diversity of options, however, there are two main categories that statistics for behavioral and social sciences can be divided into: Descriptive Statistics and Inferential Statistics.

Descriptive statistics are procedures and results that organize or summarize data for individual variables which are not generalized to populations on their own. Descriptive statistics focus on data univariately. They are named for the fact that they are used to simply describe the raw data through summaries. Thus, when descriptive statistics are used, data from variables are summarized in ways that are too limited to draw conclusions about their possible connections to each other or the likelihood that they represent truths about the populations from which their samples were drawn. Some example of descriptive statistics include means, medians, modes (which we will cover in detail in Chapter 3), ranges, standard deviations, and variances (which we will cover in detail in Chapter 4).

Inferential statistics are procedures and results that allow generalizations to be made about what is likely true of the populations from which the sample data were drawn. To **generalize** in statistics means to apply findings from sample data to the populations from which they were drawn. Another way to say this is that generalizing means inferring something about a population from sample data. When doing so, observations are being used to estimate all that existed to be observed. These are just technical ways of saying that generalizing means using sample data to estimate population data. In keeping, the term "inferential" is used to reflect the fact that truths about populations are being inferred from sample data.

Inferential techniques can be used to make inferences about individual variables for the population or the connections between or among variables in the population. Inferential techniques do this by building on various descriptive statistics (such as the mean and standard deviation) to compute probabilities about which summaries or patterns from variables measured in the samples may apply to those of the population. Essentially, inferential statistics are procedures used to estimate the likelihood that summaries and patterns in data from samples represent truths about their populations. These techniques can be univariate or multivariate. Most of the inferential statistics we will cover in this book will be bivariate in nature.

1.5: Statistics and Science

Statistics are important to many fields, including social, behavioral, and health sciences such as psychology, sociology, nursing, and education to name just a few. They are also used in many areas of the physical sciences. This is because of their important position in the scientific method.

What distinguishes sciences from other fields is their empirical nature and use of the scientific method. To be empirical is to be evidence-based or evidence-focused. Statistics is an empirical field focused on the use of data which is often used as an important step in the scientific method. The scientific method is a method of inquiry in which questions, assumptions, and/or hypotheses are tested through the collection and analysis of data before conclusions are drawn. In keeping, the use of statistics is an important part of research in a diversity of fields.

Statistics is so central to the field of psychology, for example, that the standard format for research articles in psychology includes a section detailing from whom and how data were collected, followed by a section summarizing how the data were analyzed and what the results of the analyses were, and, finally, a section discussing the implications of the findings. These sections are titled as Method, Results, and Discussion, respectively, when following the recommended format known as American Psychological Association (APA) format. The focus of this book is the analysis step of research and how it is applied. Therefore, a review of the proper way to report results in APA-format will be included in several chapters throughout this book. This will provide us some of the necessary skills to read the results from, and participate in the process of, science.

Reading Review 1.3

1. When inferential statistics are used, the goal is to _________________ results from samples to the populations from which they were drawn.
2. ________________ statistics include calculations such as the mean, median, and mode to summarize data from samples univariately.

1.6: How to Succeed in Statistics

Statistics can feel challenging at times, especially for those who are new to it. However, there are some things you should keep in mind that can help you.

First, focus on organization. There will be lots of numbers, qualitative responses, and symbols throughout your work in statistics which can get confusing and, at times, make it hard for you to identify the exact piece of information you are looking for. However, if you carefully organize the information and steps, it will make finding the information more efficient and less stressful. Imagine that there are 1,000 books about statistics all mixed up in a pile and you have to find one called “Why I Love Statistics” by Dr. Peter to do a book report. Looking for that book in the mess can be overwhelming and you might accidentally grab “Why I Loathe Statistics” by Dr. Pester instead to use for your book report. This is like having disorganized data and notes. Now imagine that instead of a pile, there are 1,000 books about statistics all nicely organized on shelves by title. You can quickly get to the book by Dr. Peter you want and even see it next to the similar book you don’t want by Dr. Pester. You can see clearly which one is the better match to what you are looking for and confidently get right to work on your book report. Let’s take this analogy a step further. Imagine you read the book and put it back in its appropriate place on the organized shelf. Then, three months later, you find out you are going to have a quiz on some topics from that book so you want to review it again. You can quickly locate it again and start refreshing your memory. Having organized and clear data and notes will benefit you as you progress through statistics. It will make it easier to follow your work and find what you need in the moment and be able to quickly return to the right information later when it is time to review.

Second, remember that the order of operations is an important process that is here to make our work easier. We will use order of operations in many places in this book. Just as organization of data and notes helps us, order of operations is a formal way of specifying the order in which the steps of formulas must be completed. This ensures that we always know what to do next and don’t get stymied in the middle of the steps to a complex formula. For a review of the order of operations, see Appendix B).

Third, go back to your foundational math skills when you get confused, frustrated, or overwhelmed. The level of statistics covered in this book uses combinations of only six basic mathematical operations: adding, subtracting, multiplying, dividing, squaring, and square rooting. That’s it. Though we will encounter new symbols for these operations and some formulas which can look quite complex, they are all just versions of the same six basic mathematical operations. For a review basic mathematical operations, see Appendix A). Fourth, consider what statistics can offer you. Statistics is an applied field focused on understanding aspects of the real world. It is used in many disciplines and careers so nearly everyone can find applicability to their academic and occupational pursuits. There are several fields and careers where the use of statistics is obvious such as mathematics, economics, biostatistics, actuarial science, and research more broadly. In addition, health statistics are important to nurses and doctors, test theory is important to educators, developmental modeling is important to child and adolescent development professionals, experimental analytics are important to the sciences, business, health, and education, and knowing how to read and explain statistics is important for journalists, meteorologists, and financial planners. In addition, several roles in the technology and engineering sectors such as user experience (UX) researcher, data scientist, data engineer, machine learning specialist, and engineering statistician require knowledge in statistics. Further, those who pursue careers as content creators are now provided with a variety of analytics that they can learn to use to potentially enhance their success. Thus, statistical skills have broad (and potentially highly valuable) application.

Fifth, statistics can help us to circumvent our biases and make better decisions. Statistics are all around us. They are used to make decisions about us but can also be harnessed to improve decisions made by us. Each of us has information available to us that is at least partially limited to or by our experiences. This in itself is not a problem; it is simply a description of an aspect of human reality. Unfortunately, however, our limited experiences can lead us to biased conclusions and poorer decision-making. Data which capture experiences beyond our own can allow us to see a bigger picture and statistics can allow us to see patterns that aren’t readily apparent through our day-to-day experiences. Thus, when used properly and thoughtfully, statistics can help us circumvent our own biases, become more knowledgeable about our worlds, and become better-informed decision makers.

Sixth, and related, take clear notes including examples to help you as you study statistics. One of the beauties of working in statistics is that we can use real world examples to help us value and think through problems. Thinking about real-world topics or questions can help you understand what the numbers represent and why the formulas include the steps that they do. This can keep you connected to the “why” of statistics at all levels: Why measure the data this way? Why do this specific mathematical operation? Why analyze the data in this way? Why are these patterns appearing in the data and what does this tell us about the world? Why bother with statistics? We will begin to answer these questions in this book.

CHAPTER OVERVIEW

2: Summarizing Data Visually

2.1: Visually Summarizing Data Univariately

To do research we must start with a question and follow this by collecting or acquiring data to help us answer that question. It is important that we use valid and reliable measures and that our data are collected in a way that allows us to do the type of analysis we wish to do. Therefore, as researchers in the behavioral sciences (such as psychology, education, and sociology) we start with a theory, generate our hypotheses (testable expectations based on previous theory, findings, or arguments), and collect and analyze data that will allow us to test our hypotheses. Statistics are used to summarize and analyze data as part of this process. One way to summarize data is to display it visually using tables or graphs. Tables and graphs simplify the data and can be a good place to start one's analyses. Therefore, the focus of this chapter will be how to summarize and read data visually using tables and graphs.

2.2: How to Read a Data Set

Variables are often measured many times in an attempt to better understand them. The sample of cases is gathered and then data are gathered about variables from those cases. The word **case** refers to an individual or on instance of a thing which is being studied. Variables are measured within cases. A **score** is a number yielded for a case. All cases may be looked at together when trying to understand the variable. However, it can be overwhelming to try to understand the variable by looking at all the cases one by one; therefore, statisticians and mathematicians use a variety of techniques to summarize score from all of the cases in a data set.

Before we can begin to use data, we must understand how to read a data set. Data can be collected and organized into a data set a few ways but the most common way to construct a data set is to put the names of variables as column headers with the data for each variable entered in rows below their appropriate headers, as shown in Data Set 2.1 below.

Table 1 Raw Data Set 2.1

ID Number	Major	Job	Years of Experience	Annual Salary	Morale
1	Statistics	UX Researcher	4	184,600	8
2	Statistics	Science Journalist	6	74,490	5
3	Statistics	Sr. Research Fellow	7	190,000	7
4	Statistics	Research Analyst	4	138,000	6
5	Statistics	Biostatistician	5	167,000	4
6	Statistics	Psychometrician	6	120,000	3
7	Statistics	Data Scientist	5	110,240	6
8	Statistics	Research Assistant	5	57,750	7
9	Statistics	Biostatistician	3	102,000	6
10	Statistics	Research Assistant	3	48,200	3
11	Statistics	Data Scientist	4	65,500	5
12	Statistics	Research Analyst	2	74,500	9
13	Statistics	Data Scientist	7	181,000	8
14	Statistics	UX Researcher	6	123,700	10
15	Statistics	Research Analyst	5	65,000	4
16	Statistics	UX Researcher	1	88,950	5
17	Statistics	Research Assistant	1	34,000	7
18	Statistics	Research Analyst	2	52,680	2

In Data Set 2.1, there are 5 columns of data. The data set shown includes raw scores. These can also be referred to as *raw data*. The term **raw scores** refers to data as they were collected and before they have been transformed, summarized, or analyzed using any formulas. Let's take a look at each column (i.e. the vertical sections of the table) of raw data in dataset 2.1 to understand it.

The first column includes ID Numbers. This is a nominal variable used to organize the cases. Each case in Data Set 2.1 was given an ID number in place of a name. For Data Set 2.1, each case has been named with sequential case numbers starting at 1. However, random numbers or letters could also be used to name cases. This can be done to help organize data without using identifying information such as a person's real name, IP address, or Social Security Number. When data are specifically presented this way to protect the anonymity of participants, the data are referred to as **de-identified data.** It is important to note that though ID Number is a variable, it is not a test variable. This means it is simply used to organize things but that it would not be analyzed or used to test hypotheses.

Each row of the data set (i.e. the horizontal sections of the table) contains the data for an individual case. Each case is a member of the sample. There are 18 rows so the sample size is 18. The symbol for sample size is n. Thus, we can summarize the sample size by writing n = 18. Sample size is always shown as a whole number because we cannot have half a case or half a person and, therefore, specifying to the hundredths decimal place does not add any information we wouldn't already know.

The second column of data is titled "Major." Major refers to the focus of someone's college degree. This was measured qualitatively on the nominal scale of measurement. However, if we read through the data for Major we will notice that, despite the fact that there are many possible majors, everyone in the sample majored in statistics. Therefore, Major is a constant in this data set and not a variable. It is best used to describe the sample and to whom limited generalizations should be made. Specifically, the fact that major did not vary means the sample of individuals is all persons who studied statistics and, therefore, any other findings or summaries from this sample would best be used to understand the population of persons who studied statistics as well rather than students of all majors.

The last four columns include data for variables which can be summarized or used to test hypotheses. The third column of data is titled "Job." Job refers to each individual's job title. This was measured qualitatively on the nominal scale of measurement. Job is a variable in this data set because not all persons had the same job title. The fourth column of data is titled "Years of Experience." This refers to the number of years of job-relevant experience someone has, rounded to the year. Years of Experience was measured quantitatively on the ratio scale of measurement. The fifth column of data is titled "Annual Salary" and reports each person's income for the year rounded to the dollar. Annual Salary, therefore, was also measured quantitatively on the ratio scale of measurement. Finally, the sixth column of data reports each person's self-reported level of "Morale" on an 11-point scale from 0 to 10 where higher values indicate greater morale. When psychological and emotional variables like morale are measured in this way, they are generally treated as interval scales. This is because the intervals between values are treated as even but the 0 is not known to represent a complete absence of the variable. It is worth noting that whether this kind of measure is truly interval or not has been debated. When certain conditions are met, such as when an 11-point scale is used, it can be appropriate to view and treat variables such as Morale as interval. Though this debate is beyond the scope of this book, those interested in learning more can start by reading Wu and Leung (2017). For this chapter, we will presume the conditions are met and treat Morale as a quantitative variable measured on the interval scale.

It is important to look through a data set to identify which columns include variables and to then assess whether each variable is quantitative or qualitative before identifying which scale of measurement was used for each variable. This is because there are different ways statisticians can summarize and present data and results that are best suited for, or can only be used with, certain kinds of data. Some ways of summarizing data can only be used for quantitative data, some for qualitative data, and some require that both quantitative and qualitative data are used together. For this chapter, we will focus on data univariately. Therefore, in this chapter we will learn how to create and read tables and graphs, some of which are used to summarize one quantitative variable at a time and others which can be used to summarize one qualitative variable at a time. Some of the graph can be extended to include more than one variable. When that is possible, it will be noted.

2.3: The Value of Visual Summaries

Visual summaries such as tables and graphs can reduce how overwhelming it is to look at all the raw data. Let's use the variable salary for example. Take a moment to look at the data in the column for the variable Annual Salary in Data Set 2.1. You might have to look at the data several times to start to get a sense of how high, low, common, or rare the various salaries were. Though it is possible to get a sense of the variable and how scores are distributed from Data Set 2.1, it is harder than it needs to be. For comparison, take a look at Salary Reordered Data Set 2.1 below (see Table 2). You will likely find that it is easier to get a sense of the data when they are reordered for the focal variable; reordering the scores from highest to lowest reduces the effort it takes to understand the data. The amount of mental effort we exert to understand something is known as *cognitive load*. If you find it faster and easier to get a sense of the data when they are in order rather than in disorder, it is because the cognitive load has been reduced. Notice that when salary was reordered so that salaries were presented in order form highest at the top to lowest at the bottom, the data for the other variables were rearranged as well. Recall that each row represents a case. Therefore, when a value for one variable gets moved, the whole row moves with it. Compare the data for the case whose ID number is 1. They were the top row in Raw Data Set 2.1 but are the second row in Salary Ordered Data Set 2.1; despite this movement, however, all their data stayed the same. The data for this case still indicate the person is a UX researcher with 4 years of experience who makes $184,600 annually.

Table 2 Salary Reordered Data Set 2.1

ID Number	Major	Job	Years of Experience	Annual Salary	Morale
3	Statistics	Sr. Research Fellow	7	190,000	8
1	Statistics	UX Researcher	4	184,600	7
13	Statistics	Data Scientist	7	181,000	8
5	Statistics	Biostatistician	5	167,000	4
4	Statistics	Research Analyst	4	138,000	6
14	Statistics	UX Researcher	6	123,700	10
6	Statistics	Psychometrician	6	120,000	3
7	Statistics	Data Scientist	5	110,240	6
9	Statistics	Biostatistician	3	102,000	6
16	Statistics	UX Researcher	1	88,950	5
12	Statistics	Research Analyst	2	74,500	9
2	Statistics	Science Journalist	6	74,490	5
11	Statistics	Data Scientist	4	65,500	5
15	Statistics	Research Analyst	5	65,000	4
8	Statistics	Research Assistant	5	57,750	7
18	Statistics	Research Analyst	2	52,680	2
10	Statistics	Research Assistant	3	48,200	3
17	Statistics	Research Assistant	1	34,000	7

Reordering the data is especially helpful when trying to understand quantitative data because it allows the statistician to easily see the highest and lowest values. For example, when the data are reordered for annual salary, it is easy to quickly discern that the lowest salary was $34,000 and the highest was $190,000. That's quite a large range. We could do the same by reordering data based on years of experience to quickly and easily see that the least experience in the data set was 1 year and the highest was 7 years (see Table 3).

Table 3 Experience Reordered Data Set 2.1

ID Number	Major	Job	Years of Experience	Annual Salary	Morale
3	Statistics	Sr. Research Fellow	7	190,000	7
13	Statistics	Data Scientist	7	181,000	8
14	Statistics	UX Researcher	6	123,700	10
6	Statistics	Psychometrician	6	120,000	3
2	Statistics	Science Journalist	6	74,490	5
5	Statistics	Biostatistician	5	167,000	4
7	Statistics	Data Scientist	5	110,240	6
15	Statistics	Research Analyst	5	65,000	4
8	Statistics	Research Assistant	5	57,750	7
1	Statistics	UX Researcher	4	184,600	8
4	Statistics	Research Analyst	4	138,000	6
11	Statistics	Data Scientist	4	65,500	5
9	Statistics	Biostatistician	3	102,000	6
10	Statistics	Research Assistant	3	48,200	3
12	Statistics	Research Analyst	2	74,500	9
18	Statistics	Research Analyst	2	52,680	2
16	Statistics	UX Researcher	1	88,950	5
17	Statistics	Research Assistant	1	34,000	7

Statistics can be complicated and sometimes feel difficult. However, we can use visual displays to ease the mental effort it takes to understand the data even further. Therefore, statisticians follow some basic strategies and rules for organizing, summarizing, and presenting data that are meant to reduce cognitive load, thereby making things easier to understand. So far we have seen the efficiency gained simply putting quantitative values in descending order. When statisticians are interested in summarizing how frequently various data occurred, however, reordering is not sufficient. Therefore, specific tables have been designed and used to make it easy to deduce the frequency with which various raw scores occurred.

2.4: Frequency Distribution Tables

There are a variety of tables that can be used to summarize the frequency with which scores were observed for a quantitative variable. The most basic version of this is simply referred to as a frequency distribution.

Creating Frequency Distributions using Scores

A frequency distribution is a table used to summarize a quantitative variable by showing how frequently each score occurred. A frequency distribution has two columns. When the range of values is somewhat small (i.e. generally the difference between the highest score and the lowest score is 20 or less), the first column is titled *Score*. The score column shows the various scores which could have occurred starting from the highest which was observed (top row of data) to the lowest which was observed (bottom row of data). The second column is titled *Frequency*. To fill in frequency column, the statistician counts up how many times each score was observed in the data. Then, they put the count in the corresponding row. Take a look at Table 4 below. This is a frequency distribution for the variable Years of Experience built using the scores from Data Set 2.1.

Table 4 Frequency of Years of Experience (n = 18)

Score	Frequency
7	2
6	3
5	4
4	3
3	2
2	2
1	2

Interpretation

Let's review how to read the table before going over the rules used to create it. The title tells us the variable being summarized and the sample size. The variable is Years of Experience and the sample size is 18. The score column shows all the possible scores that could have occurred ranging from the highest which was observed down to the lowest which was observed. We can see that experience ranged from 1 to 7 years. The frequency column tells us how frequently each score occurred. In this table, the frequencies refer to how many cases in the data set were people who had 1, 2, 3, 4, 5, 6, or 7 years of experience. We can see that the most frequent number of years of experience was 5 years because 4 cases were of people with this many years of experience. The least common years of experience reported were 7, 3, 2, and 1 because each of these amounts of experience were reported by two people. We also know that no one in the dataset reported having 0 years of experience or more than 7 years of experience because if they had, we would have rows for those scores in the table. If you sum the values in the frequency column it will be equal to the sample size. This is because the frequency column shows how many cases in the sample had each score and, thus, was created by simply dividing up the sample.

Construction

To construct a frequency distribution a statistician starts by organizing the data to get a sense of the range. We did this with the data for Years of Experience in Table 3. The statistician must then identify the highest score, the lowest score, and the precision with which they were measured and/or rounded. Years of Experience was measured and shown to the year in intervals of 1 year. The statistician then uses this information to construct the score column. Though it varies for different fields and circumstances, it is common in many behavioral and social sciences to create the score column going from the lowest score at the bottom toward the highest score at the top. Therefore, the score column for Table 4 was constructed counting up from lower scores (starting at 1) to higher scores (ending at 7) in increments of 1. However, you may also see tables which are constructed from lowest on the top to highest at the bottom when this is the better fit for the field or data.

When a score column is constructed, no rows are needed for values outside the range within which data were observed. This makes it easy for anyone reading the table to quickly deduce the range and know that scores beyond it were not observed. If we want to know how many people had 0 years of experience or 10 years of experience in Table 4, for example, we can quickly deduce the answer to both is 0 because these are outside the range shown and, thus, outside the range observed. However, a row for any score inside the observed range is retained in the table even it if has a frequency of 0. Suppose that no one had 3 years of experience but that all other data were the same for Table 4. In this case, we would simply put a 0 in the frequency column to the right of the score of 3 rather than deleting the row corresponding to this score.

Ideally, a frequency distribution should have between 10 and 20 rows, however, there are times when it is appropriate to have more than 20 rows or fewer than 10 rows. When there are more than 20 rows the table can be providing too much detail without summarizing enough. It also increases the cognitive load for the reader when there are many rows. When there are fewer than 10 rows, it sometimes means that the data are being over-simplified and too little detail is being shown. However, if the data have a small range, we sometimes must use fewer than 10 rows. Take a look again at Table 4. There are only 7 rows and those are sufficient to show all the detail needed and possible for the variable Years of Experience because the range was small. Therefore, it is appropriate to have a small table with only 7 rows for these data.

Creating Frequency Distributions using Intervals

However, there are times when the range of values for a variable is so large that all scores could not be listed in about 20 or fewer rows. When this occurs, statisticians create a frequency distribution using an interval column in place of a score column. Intervals break the range up into useful segments which summarize and organize the data. To enhance clarity and consistency, there are rules that should be followed when creating intervals. First, the intervals should all be the same size. Second, the intervals should be intuitive. By intuitive here we mean that the intervals should be in amounts that are easy to count in or understand such as ones (0, 1, 2, 3, 4, 5, and so on), twos (0, 2, 4, 6, 8, 10 and so on), fives (0, 5, 10, 15, 20, 25 and so on), tens (0, 10, 20, 30, 40, 50 and so on) or hundreds (0, 100, 200, 300, 400, 500 and so on). Third, the intervals must be mutually exclusive. This means the intervals cannot overlap so that each score fits into only one interval. Finally, it is best to have approximately 10 to 20 intervals as appropriate for the data. This last rule must be a bit flexible to allow the other three rules to be followed.

A variable such as Annual Salary from Data Set 2.1 necessitates the use of intervals. This is because values were measured to the dollar and ranged from 34,000 to 190,000. If each dollar amount from 34,000 to 190,000 got its own row, the table would have 156,001 rows. This would be overwhelming both to create and to read and wouldn't meet the primary goal of making the data easier to understand through a summary. No one wants that. Instead, intervals can be created to organize the data more efficiently.

Interval Construction

Creating intervals takes a bit of practice and thought but a good place to start is by finding the range for the variable. The number of dollar amounts starting from the highest salary down to and including the lowest salary, also known as the inclusive range, is 156,001. If we want 10-20 rows, we need to figure out how to divide this range intuitively and evenly. A good strategy is to divide the range by 10 to see the approximate interval size it would take to have 10 rows, then divide the range by 20 to see the approximate interval size it would take to have 20 rows. If we do this we will find the approximate interval size to get 10 rows is 15,600 and for 20 rows is 7,800. Neither of these are intuitive but we can choose a value somewhere between them that is intuitive such as 10,000. Intervals of 10,000 are easy to count in and understand and would cause our table to have about 17 rows to accommodate the full range of data. That is between 10 and 20 rows so it meets two of our criteria. Last, we need to create the intervals of 10,000 which are mutually exclusive. An easy and advisable place to start creating intervals for ratio level data such as salary is 0. Therefore, we can start by creating our lowest interval going from 0-9,999. This is an interval with 10,000 dollar amounts in it because 0 counts as the first value (making 1 the second value, 2 the third value and so on until we get to 9,999 as the 10,000th value).

Let's look at the construction process for creating a frequency distribution for Annual Salary. We started with 0-9,999. We continue to the interval 190,000-199,999 because this top interval will include the highest salary observed in the data. This causes us to have three intervals at the bottom that we do not need because the lowest income observed was 34,000. Therefore, we can remove those three before we create our frequency column. Drafting those bottom intervals and removing them can help folks know where to start their intervals so they won't be tempted to start the interval at the lowest value of 34,000 which could cause the bottom interval to go from 34,000-43,999 which is less intuitive and, thus, creates more work for the reader. However, some people can skip the steps of drafting and removing the bottom rows if they can remember to start their intervals at intuitive numbers (i.e. start at 30,000 rather than 34,000 for this table). Drafting and removing the bottom rows or simply starting at 30,000 are equally appropriate ways to create the interval column so it is best to use the method that makes the most sense to you.

Table 5 Interval Colum for Annual Salary

Interval

Interval
190,000-199,999 180,000-189,999 170,000-179,999 160,000-169,999 150,000-159,999 140,000-159,999 130,000-139,999 120,000-129,999 110,000-119,999 100,000-109,999 90,000-99,999 80,000-89,999 70,000-79,999 60,000-69,999 50,000-59,999 40,000-49,999 30,000-39,999 ~~20,000-29,999~~ ~~10,000-19,999~~ ~~0-9,999~~

Next, we count how many salaries fell into each interval. This is easiest to do by first putting the data for the variable in order and then counting the occurrences. Therefore, we can look at Table 2 and begin counting. One raw score for salary was between 30,000 and 39,000 so the frequency for this interval was 1. One raw score was between 40,000 and 49,999 so the frequency for this interval was also 1. However, two raw scores were between 50,000 and 59,999 so the frequency for this interval is 2. We continue counting in this way until we have identified and filled in the frequencies for all intervals as shown in Table 6. Once all the frequencies are filled in it is good practice to sum them to make sure they are equal to the sample size. If we add the frequencies in Table 6 we get 18 which is equal to the sample size so no errors are readily apparent.

Table 6 Frequency Distribution for Annual Salary (n = 18)

Interval	Frequency
190,000-199,999	1
180,000-189,999	2
170,000-179,999	0
160,000-169,999	1
150,000-159,999	0
140,000-159,999	0
130,000-139,999	1
120,000-129,999	2
110,000-119,999	1
100,000-109,999	1
90,000-99,999	0
80,000-89,999	1
70,000-79,999	2

Interval	Frequency
60,000-69,999	2
50,000-59,999	2
40,000-49,999	1
30,000-39,999	1

Interpretation

Now that the frequency distribution has been created, let's take a moment to read it, paying close attention to a few thing. First, we can quickly deduce that no interval was especially common or uncommon relative to other intervals as there were between 0 and 2 cases for all 17 intervals shown. Thus, the incomes were spread out somewhat evenly with half the sample (9 cases) earning 100,000 or more and half (the other 9 cases) earning 89,999 or less. It is also quick and easy for us to identify how many salaries in the sample were in any given interval. For example, if you want to quickly know how many people earned between 50,000 and 59,999, you can look down to that row and easily see the count was 2. Compare this to trying to quickly identify how many salaries were between 50,000 and 59,999 in the raw data show in Table 1. It is going to take more time and effort to answer the question from the raw data.

Retaining Empty Rows

Some may wonder why we don't remove the intervals from the table with a frequency of 0 to make the table look even simpler. The irony is that removing those intervals makes the table shorter yet increases the cognitive load. Let's take a look to see why. Try to find the frequency with which salaries between 90,000 and 99,000 were observed in the improperly constructed version shown in the right side Table 7. Once you have your answer try it again using the properly constructed version shown in the left side Table 7.

Table 7 Proper and Improper Interval Columns for Annual Salary

Proper Construction		Improper Construction	
Interval	**Frequency**	**Interval**	**Frequency**
190,000-199,999	1	190,000-199,999	1
180,000-189,999	2	180,000-189,999	2
170,000-179,999	0	160,000-169,999	1
160,000-169,999	1	130,000-139,999	1
150,000-159,999	0	120,000-129,999	2
140,000-159,999	0	110,000-119,999	1
130,000-139,999	1	100,000-109,999	1
120,000-129,999	2	80,000-89,999	1
110,000-119,999	1	70,000-79,999	2
100,000-109,999	1	60,000-69,999	2
90,000-99,999	0	50,000-59,999	2
80,000-89,999	1	40,000-49,999	1
70,000-79,999	2	30,000-39,999	1
60,000-69,999	2		
50,000-59,999	2		
40,000-49,999	1		

Proper Construction		Improper Construction	
30,000-39,999	1		

It is easier to confidently deduce that the frequency of salaries between 90,000 and 99,999 in the sample was 0 using the properly constructed table than in the improperly constructed table. Next, take a look down the frequency column in each table. In the properly constructed table, the pattern is clear that some intervals occurred once or twice while others did not occur at all. We don't need to look at the interval column to know if any intervals had a frequency of 0; instead we only need to look at the frequency column to see which intervals had a frequency of 0. However, if we look down the frequency column of the improperly constructed version of the table, we cannot easily tell which, if any, intervals had a frequency of 0. Instead, we have to look at the interval column and try to figure out which intervals are missing. If an interval is missing we would assume that it was because the frequency was 0 but may not feel as confident. We might wonder if the person who created the table made a mistake and left out an interval. To check this we could add all the frequencies to see it if is equal to the sample size. If so, we can be fairly confident that the missing interval isn't a mistake but rather is meant to indicate the frequency for that interval was 0. We *could* do all this extra work but why? This is a lot more work than simply looking at the row for the interval in the properly constructed table to see if the frequency was 0. Thus, the reason intervals within the range with a frequency of 0 are retained is to meet the goal of making the data easier to understand and summarize.

Reading Review 2.1

Two of the tables below have construction errors and one is properly constructed. Identify which two have errors and specify the nature of the errors.

Data Set A, Daily Inches of Rainfall: 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 3 3 3 3 3 4 4 4 7 7

Table A Frequency of Daily Inches of Rainfall (n = 31)

Score	Frequency
7	2
6	0
5	0
4	3
3	5
2	4
1	8
0	9

Data Set B, Age in Years: 2 4 6 6 7 9 10 11 11 14 15 18 18 18 19 20 22 23 23 24 26 26 28 33

Table B Age in Years (n = 24)

Interval	Frequency
20-39	9
0-19	15

Data Set C, Length in Inches: 2 2 3 4 5 7 8 9 9 11 12 12

Table C Length in Inches (n = 12)

Score	Frequency
12	2
11	1

Score	Frequency
9	2
8	1
7	1
5	1
4	1
3	1
2	2

Cumulative Frequency Distributions

A third column can be added to a frequency distribution to create a table known as a cumulative frequency distribution. **Cumulative frequency distributions** are used to summarize the frequencies and cumulative frequencies of ordered scores for quantitative variables. The third column is added to the right of the frequency column and is given the title "Cumulative Frequency." This cumulative column counts up from the bottom row to tell how many cases occur at or below each row.

Let's take a look at an example. Table 8 is a cumulative frequency distribution for the variable Annual Salary. The first two columns are the same as those used in the frequency distribution for these data but a third column has been added. Starting from the bottom and counting up, the cumulative frequency column summarizes how many people were at or below each salary interval. In this example, starting from the bottom we can see that 1 case for salary was at or

below 39,999, 2 cases were at or below 49,999, 4 cases were at or below 59,999 and so on until we reach the highest interval. The total sample size appears as the top cumulative frequency because all scores are at or below the highest score or interval observed.

Table 8 Cumulative Frequency Distribution for Annual Salary (n = 18)

Interval	Frequency	Cumulative Frequency
190,000-199,999	1	18
180,000-189,999	2	17
170,000-179,999	0	15
160,000-169,999	1	15
150,000-159,999	0	14
140,000-159,999	0	14
130,000-139,999	1	14
120,000-129,999	2	13
110,000-119,999	1	11
100,000-109,999	1	10
90,000-99,999	0	9
80,000-89,999	1	9
70,000-79,999	2	8
60,000-69,999	2	6
50,000-59,999	2	4
40,000-49,999	1	2
30,000-39,999	1	1

Let's review it one more time to better understand what the third column shows us. Notice that the bottom right shows a cumulative frequency of 1 because there is only 1 case at the interval of 30,000-39,999 or below (since there are no data below). Notice also that the second row from the bottom has a cumulative frequency of 2 because it adds 1 (the number of cases in the second interval from the bottom) to 1 (the total number of cases that occur below the second interval from the bottom). Cases are added as we move up the cumulative column so that we know how many cases are at or below each interval. This allows us identify specific thresholds easily.

Relative Frequency Distributions

There are other options for creating a third column, each of which makes a different kind of frequency distribution. One option is to add a third column to make a relative frequency distribution. **Relative frequency distributions** (also known as proportional frequency distributions) are used to summarize the relative frequencies of scores or intervals using percentages. The third column is given the title "Relative Frequency." This column what percentage of the raw scores for the variable are represented in each row.

Let's take a look at an example. Table 9 is a relative frequency distribution for the variable Annual Salary. The first two columns are the same as those used in the frequency distribution for these data but a third column has been added. Each row of the relative frequency column is computed by dividing the frequency for each row by the total sample size and then multiplying by 100 to calculate the relative percentage. For example, the frequency for the top interval of 190,000-199,999 is 1. The sample size is 18. Thus, the percentage is computed as follows:

$(1 \div 18)100 = (0.555\ldots)100 = 5.56\%$

The sum of all values in the relative frequency column must be 100% (give or take any rounding error introduced when reporting percentages for each interval).

Table 9 Relative Frequency Distribution for Annual Salary (n = 18)

Interval	Frequency	Relative Frequency
190,000-199,999	1	5.56%
180,000-189,999	2	11.11%
170,000-179,999	0	0.00%
160,000-169,999	1	5.56%
150,000-159,999	0	0.00%
140,000-159,999	0	0.00%
130,000-139,999	1	5.56%
120,000-129,999	2	11.11%
110,000-119,999	1	5.56%
100,000-109,999	1	5.56%
90,000-99,999	0	0.00%
80,000-89,999	1	5.56%
70,000-79,999	2	11.11%
60,000-69,999	2	11.11%
50,000-59,999	2	11.11%
40,000-49,999	1	5.56%
30,000-39,999	1	5.56%

Cumulative Percentage Distributions

A cumulative frequency distribution can also be made using percentages by either, creating a third column that reports the cumulative relative proportions or by adding a cumulative column to a relative frequency distribution. Thus, **cumulative percentage distributions** are often created as relative frequency distributions with a fourth column showing the cumulative percentages as shown in Table 10. Just like a cumulative frequency distribution table, the cumulative column is made by summing computations starting from the bottom row and working up. The sum of all values in the relative frequency column must be 100% (give or take any rounding error introduced when reporting percentages for each interval). To help avoid any such rounding error, it is best practice to find the cumulative frequency for each row, to divide that by the sample size, and then multiply by 100 to get the cumulative percentage rather than adding the relative frequencies which may have rounding error. For example, the cumulative frequency for the interval of 50,000-59,999 was 4. When 4 is divided by the sample size of 18 the result is 0.222… When this is multiplied by 100 to translate from a decimal to a percentage and then rounded to the hundredths place the result is 22.22%. The same procedure was used for each row to create the Cumulative Percentage column of Table 10.

Table 10 Cumulative Percentage Distribution for Annual Salary (n = 18)

Interval	Frequency	Relative Frequency	Cumulative Percentage
190,000-199,999	1	5.56%	100.00%
180,000-189,999	2	11.11%	94.44%
170,000-179,999	0	0.00%	83.33%
160,000-169,999	1	5.56%	83.33%
150,000-159,999	0	0.00%	77.78%
140,000-159,999	0	0.00%	77.78%
130,000-139,999	1	5.56%	77.78%
120,000-129,999	2	11.11%	72.22%
110,000-119,999	1	5.56%	61.11%
100,000-109,999	1	5.56%	55.56%
90,000-99,999	0	0.00%	50.00%
80,000-89,999	1	5.56%	50.00%
70,000-79,999	2	11.11%	44.44%
60,000-69,999	2	11.11%	33.33%
50,000-59,999	2	11.11%	22.22%
40,000-49,999	1	5.56%	11.11%
30,000-39,999	1	5.56%	5.56%

Percentile Rank Distributions

You may also see or need a table that reports the percentile rank of each score or interval. A **percentile rank distributions** can be created as relative frequency distributions with a fourth column showing the percentile rank (i.e. the percentage of scores below) for each score or interval as shown in Table 11. A percentile ranks refer to the percentage of raw scores below a given score of interval. For example, 0 raw scores in the data set are below the interval of 30,000-39,999 in Data Set 2.1 Thus, the percentile rank for that row is 0.00%. However, there are two raw scores below the interval of 50,000-59,999 and, thus, the percentile rank of this interval is 11.11%. Notice that the values in the percentile rank column for Table 11 are the same as those in Table 10 just shifted up one by one row. This is because Table 10 focused only on scores below each row.

Table 11 Percentile Rank Distribution for Annual Salary (n = 18)

Interval	Frequency	Relative Frequency	Percentile Rank

Interval	Frequency	Relative Frequency	Percentile Rank
190,000-199,999	1	5.56%	94.44%
180,000-189,999	2	11.11%	83.33%
170,000-179,999	0	0.00%	83.33%
160,000-169,999	1	5.56%	77.78%
150,000-159,999	0	0.00%	77.78%
140,000-159,999	0	0.00%	77.78%
130,000-139,999	1	5.56%	72.22%
120,000-129,999	2	11.11%	61.11%
110,000-119,999	1	5.56%	55.56%
100,000-109,999	1	5.56%	50.00%
90,000-99,999	0	0.00%	50.00%
80,000-89,999	1	5.56%	44.44%
70,000-79,999	2	11.11%	33.33%
60,000-69,999	2	11.11%	22.22%
50,000-59,999	2	11.11%	11.11%
40,000-49,999	1	5.56%	5.56%
30,000-39,999	1	5.56%	0.00%

2.5: Histograms

Another way you can display your data is with a graph. There are several kinds of graphs that you can use depending upon the type of data you have and what you want to be able to see with or about those data. Let's consider the kinds of graphs we can make for the variable Years of Experience. Our goal here is simply to summarize our data from this quantitative variable univariately. A graph called a histogram works well for summarizing quantitative data and it builds on the same general ideas as the frequency distribution.

Histograms are frequency distribution graphs for quantitative data which present scores or intervals using abutted bars along the x-axis and frequencies along the y-axis. These are most appropriate for variables which are continuous and measured using an interval or ratio scale. To construct a histogram, data are first organized to identify the range that needs to be shown on the x-axis. Then the frequencies of the scores (or intervals) are counted to identify the height needed for the y-axis. Once these axes are created and labeled, vertical bars are added along the x-axis. The bar widths extend along the x-axis to the real score limits of the score of interval they represent. Essentially, this means each bar extend to the midpoint between itself and the scores to the left and right of it. Because bars extend to the real score limits, bars which represent adjacent scores or intervals abut or touch each other. This is to represent the continuous nature of many quantitative variables. The bar heights extend up corresponding with the frequency with which data occurred at the scores or within the intervals they represent.

Creating Histograms using Scores

Let's take a look at a histogram for the variable Years of Experience in Data Set 2.1 (see Figure 1). There were two cases of individuals who had 1 year of experience in the data set. We see this represented by the first bar of the graph when read from left to right. This bar is over the number 1 on the x-axis and extends to the midpoint between it and the value to the left of it (which is 0) and the value to the right of it (which is 2); thus, the bar representing 1 year of experience spans from 0.50 to 1.50 on the x-axis. To the right of it is the bar representing 2 years of experience which spans from 1.50 to 2.50 on the x axis. The bars continue in this way along the x-axis. The height of the first bar representing 1 year of experience is 2 units high on the y-axis. This is because the frequency with which data indicating 1 year of experience were observed was 2. Thus, the height of 2 indicates that two cases reported a raw score of 1 for years of experience. This continues for each bar such that their heights correspond to the frequency with which the scores they represent occurred.

Let's take note of a few other aspects of how the graph was constructed. First, there are no bars over the scores of 0 or 8. This represents an absence of data for those scores. This means that whenever we see a gap where a bar appears to be missing on a histogram, it indicates that the frequency with which that score was observed was 0. It is possible to simply eliminate the sections of the graph representing scores of 0 and 8 which would make the graph narrower. Some statisticians, however, prefer to extend the x-axis one score or interval to the left and right of the range within which data were observed to make the graph a bit easier to read. Essentially, when statisticians leave those gaps in on the left and right, they are visually confirming that the graph was not cut-off somewhere and that no bars are missing. Similarly, you can see that the y-axis extends up to a frequency of 5 despite the fact that the tallest bar only goes up to a frequency of 4. It is sometimes recommended to leave about 20-25% of extra empty space at the top of the y-axis to ensure that the tops of all the bars are clearly visible and to make the graph a bit more visually appealing.

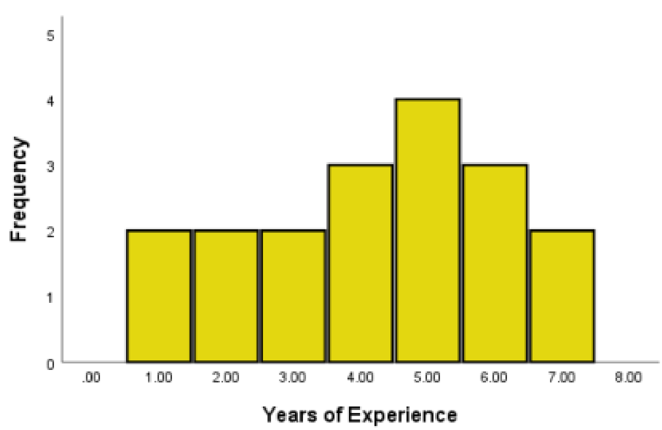


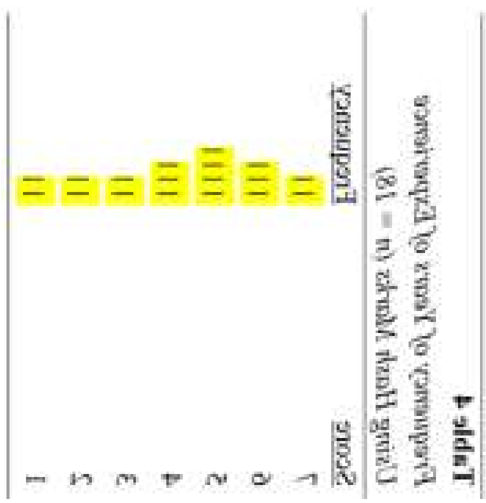
Figure 1 *Histogram for Years of Experience (n = 18)*

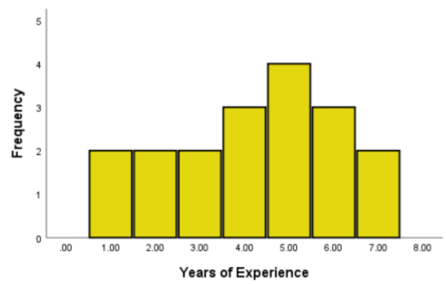
Comparing Histograms to Frequency Distributions. Histograms provide the same general information and summaries as frequency distribution tables but in graph form. Let’s compare the two for Years of Experience to make their similarities and differences clearer. We can take our frequency distribution for Years of Experience (Table 4) and tweak it slightly to use hash marks instead of numbers in the frequency column; the result goes from the version shown on the left to the one shown on the right:

Table 4

<table>
<tr><th colspan="3">Frequency of Years of Experience (n = 18)</th><th colspan="2">Frequency of Years of Experience Using Hash Marks (n = 18)</th></tr>
<tr><th>Score</th><th>Frequency</th><th></th><th>Score</th><th>Frequency</th></tr>
<tr><td>7</td><td>2</td><td></td><td>7</td><td>| |</td></tr>
<tr><td>6</td><td>3</td><td></td><td>6</td><td>| | |</td></tr>
<tr><td>5</td><td>4</td><td></td><td>5</td><td>| | | |</td></tr>
<tr><td>4</td><td>3</td><td></td><td>4</td><td>| | |</td></tr>
<tr><td>3</td><td>2</td><td></td><td>3</td><td>| |</td></tr>
<tr><td>2</td><td>2</td><td></td><td>2</td><td>| |</td></tr>
<tr><td>1</td><td>2</td><td></td><td>1</td><td>| |</td></tr>
</table>

The hash marks now use space to represent frequency horizontally the same way the bars of a histogram represent frequency vertically. If you look sideways at the hash marks in Table 4 and read them from low to high (bottom to top) you are essentially looking at a histogram. Here is the rotated and highlighted version of the frequency distribution for Years of Experience next to its histogram to illustrate this:





Notice the similarity in the highlighted hash marks in the rotated table and the bars of the histogram. Thus, either the table or the graph are useful ways of creating univariate summaries of quantitative variables but only one is typically used at any given time because using both would be redundant.

Creating Histograms using Intervals

Just like with a frequency distribution table, some data require the use of intervals when creating histograms. When the range of scores is too large to fit into approximately 20 bars or less, intervals are used instead of scores for creating the x-axis of the histogram. There are two ways the x-axis is commonly displayed when intervals are used. The first option, shown in Figure 2, is to label the boundaries between the intervals. To do so, the lower limit of each interval appears under the left (lower) edge of the bar representing it. Therefore, Figure 2 shows labels (also known as “anchors”) on the x-axis that match the lower end of each interval used in the corresponding frequency distribution table created with the same data (as shown in Table 6). Thus, we see anchors at 30,000, 40,000, 50,000, and so on.

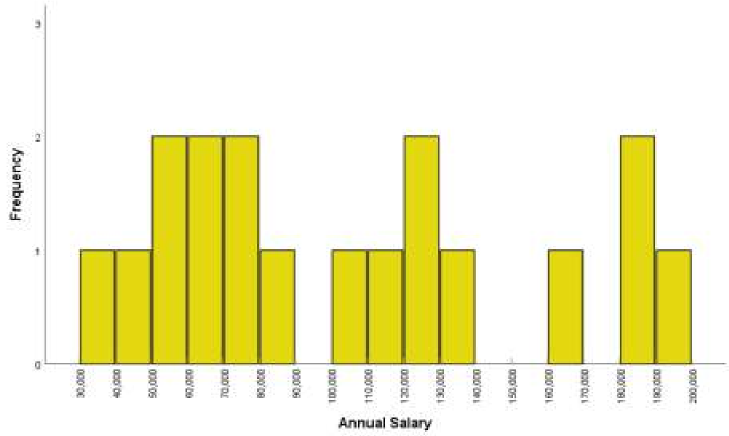


Figure 2 *Histogram for Annual Salary (n = 18)*

An alternative way to label the x-axis is by stating the midpoint of each interval centered under each bar as shown in Figure 3. When this version is used, an extra anchor to the left and right of the ones for which there are data (bars) can be used to make it easier for readers to visually locate the upper and lower limits used for each interval.

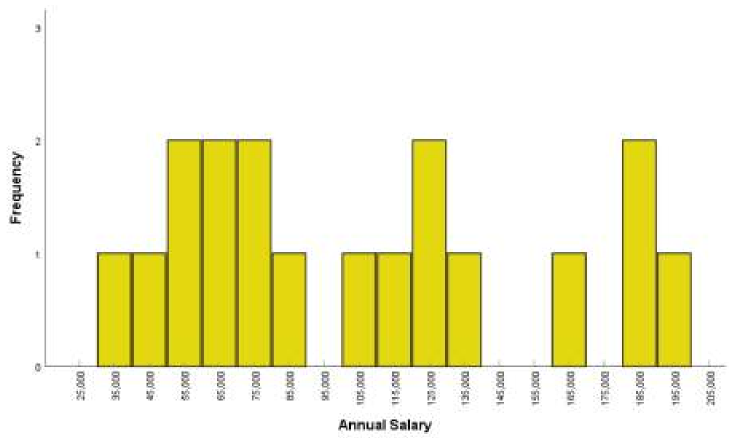


Figure 3 *Histogram for Annual Salary using Centered Anchors (n = 18)*

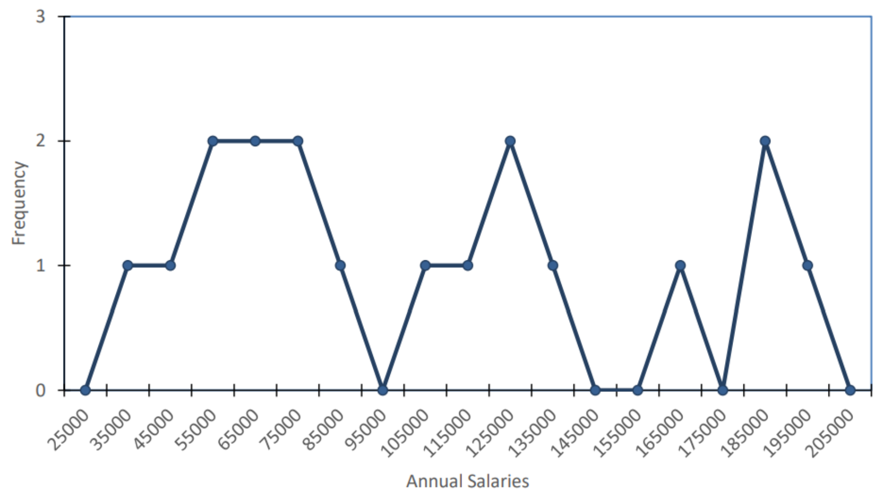
2.6: Frequency Polygons

Another option for graphing a frequency distribution of quantitative scores is a frequency polygon. Polygon means a many-sided figure. **Frequency polygons** are frequency distribution graphs for quantitative data which present scores or intervals using lines which connect to score or interval midpoints along the x-axis at varying heights which represent frequencies along the y-axis. That's a mouthful, so let's simplify: essentially, a frequency polygon replaces the bars of a histogram with a shape made by connecting the midpoints of the tops of each bar.

To construct a histogram, data are first organized to identify the range that needs to be shown on the x-axis. Then the frequencies of the scores (or intervals) are counted to identify the height needed for the y-axis. These are the same initial steps used to create a histogram. Once these axes are created and labeled, dots are placed over the center of each score or interval on the x-axis at a height corresponding to the frequency with which the corresponding score or interval was observed in the data. If the frequency of a score or interval is 0, the dot is placed on the location of the x-axis corresponding to that score or interval midpoint. Finally, dots are connected using straight lines from left to right. When creating a frequency polygon, space is provided on the left and right of the range within which scores occurred. Those two locations have frequencies of 0 and are included so that the graph can start and end by anchoring down to the x-axis.

Figure 4 shows a frequency polygon using the annual salary data from Data Set 2.1. This was constructed using intervals. If you compare it to the histogram for these data (Figure 3), you can see the similarity in heights and patterns but the difference in how the shift from interval to interval is represented visually as bars giving a somewhat more discrete appearance to the data (in the histogram) or sloping shifts which give a somewhat more continuous appearance to the data (in the polygon).

Figure 4. Frequency Polygon of Annual Salary Intervals



2.7: Bar Graphs

Univariate Bar Graphs. Frequency distributions, cumulative frequency distributions, histograms, and frequency polygons are all appropriate for continuous, quantitative data measured on interval or ratio scales but not for qualitative data nor discontinuous, quantitative data. When statisticians want to report the frequencies of different categories for qualitative data (such as college major) or discontinuous, quantitative data (such as number of children), they can use a bar graph. Though bar graphs look very similar to histograms, they can be easily distinguished by the fact that spaces are left between adjacent bars in bar graphs. This space between the bars is meant to visually represent the distinction between categories. Bar graphs are sometimes also created for discontinuous, quantitative data measured and those measured on an ordinal scale. For discontinuous, quantitative data, the gaps between the bars serve as a visual reminder that the ordered categories either are discontinuous. For ordinal data, the gaps serve as a visual reminder that the intervals between the data are not known to be the same size or distance and, thus, that the visual midpoints between them are not necessarily quantitatively equivalent differences.

Similar to histograms, bar graphs represent the diversity of data on the x-axis and the frequency with which the various observations occurred on the y-axis. Unlike histograms, however, the x axis of a bar graph shows categories of data (such as job titles or gender identities) which generally do not have a prescribed order. There are a few ways to present the categories on the x-axis but a fairly common option is to put the categories in alphabetical order when the data consist of letters, words, or phrases. Bars are then constructed centered over their anchors on the x-axis. Evenly sized gaps are left between bars to emphasize the distinction of categories. These gaps should be wide enough to be apparent but narrower than the bars representing the data. The height of the bars represents frequency, however, the y-axis is often labeled with the word *count* instead of *frequency*, though both are generally considered acceptable.

Bar graphs can be used to quickly deduce which qualitative responses occurred and which of those were more or less common than one another. Let's take a look at an example of a bar graph and how to interpret it. Job Title in Data Set 2.1 is qualitative and, thus, the counts of the various job titles can be summarized on a bar graph but not on a histogram; Figure 5 shows a bar graph for those data. There are 8 bars which means that there were 8 distinct job titles observed in the data set. Note the fact that the bars are all the same width. Adjacent bars do not abut or touch to represent the distinction between categories. Note also that the gaps separating the bars are all the same width and are narrower than the bars. This allows the bars to get the visual emphasis over the gaps. In reading the graph we glean that the most common of the 8 job titles was Research Analyst which appeared in the data set four times. We can also see that Data Scientist, Research Assistant, and UX Researcher were equally common as each appeared in the data set three times. Biostatistician appeared in the data twice and the remaining three job titles each appeared in the data once. If we sum the heights of the bars we get 18 which was the total sample size.

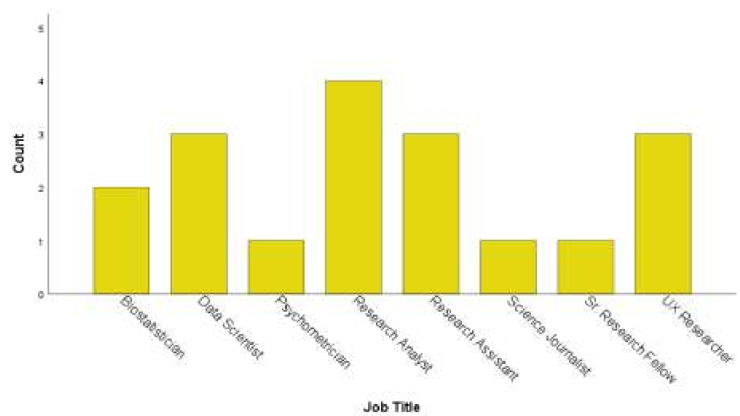


Figure 5 *Counts of Job Titles (n = 18)*

Reading Review 2.2

1. Under what conditions would an interval column be used in place of a score column in a frequency distribution?
2. What is indicated in each row of the cumulative column in a cumulative frequency distribution?
3. Why are adjacent bars shown abutting or touching in histograms?
4. What is represented by bar heights in histograms?

5. For which forms of data are bar graphs the appropriate choice over histograms? 6. What is represented by bar heights in bar graphs?

2.8: Using SPSS

Software can be used to expedite analyses and aid in the creation of tables and graphs. One that is commonly used by behavioral and social scientists and statisticians is called SPSS (which is owned by IBM). SPSS stands for Statistical Package for Social Sciences. This software can be used to quickly organize, summarize, and analyze data once data have been properly entered into the program.

Entering Data

Start by identifying how many variables you have and whether data for each are represented with numbers or letters. Then, identify whether each was measured on the nominal, ordinal, interval, or ratio scale.

Next, open the SPSS software. Click "New Dataset," then click "Open." In some versions of SPSS the word "Open" is replaced with "OK." This will create a new blank spreadsheet into which you can enter data. There are two tabs which appear towards the bottom of the spreadsheet. One is called "Variable View" which is the tab that allows you to tell the software about your variables. The other is called "Data View" which is the tab that allows you to enter your data.

Click on the Variable View tab. This tab of the spreadsheet has several columns to organize information about the variables. The first column is titled "Name." Start here and follow these steps:

1. Click the first cell of the "Name" column and enter the name of your first variables using no spaces, special characters, or symbols. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
2. Click the first cell of the column titled "Type" and then click the three dots that appear in the right side of the cell. Specify whether the data for that variable appear as numbers by selecting "Numeric" or as letters or words by selecting "String." There are other options for specific kinds of data such as dates and currency which can be selected as appropriate for the data. Then click "OK." For numeric data SPSS will automatically allow you to enter values that are up to 8 digits in length with decimals shown to the hundreds place as noted in the "Width" and "Decimal" column headers, respectively. You can edit these as needed to fit your data, though these settings will be appropriate for most quantitative variables in the behavioral sciences.
3. Click the first cell of the column titled "Label." This is where you can specify what you want the variable to be called in output, including in tables and graphs. You can use spaces or phrases here.
4. Click on the first cell of the column titled "Measure." A pulldown menu with three options will allow you to specify the scale of measurement for the variable. However, SPSS does not differentiate between interval and ratio and, instead, refers to both of these as "Scale." Select the option that best fits the data for your variable.
5. Move to the second row of the spreadsheet, starting with the cell under "Name" and repeat steps 1-4 until you have entered the information for all of your variables.

Here is what the Variable View tab would look like when created for Data Set 2.1:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ID	Numeric	8	2	Years of Experience	None	None	8	Right	Nominal
2	Major	String	10	0	Major	None	None	10	Left	Nominal
3	Job	String	20	0	Job Title	None	None	20	Left	Nominal
4	YearsOfExperience	Numeric	3	0	Years of Experience	None	None	3	Right	Scale
5	AnnualSalary	Numeric	8	2	Annual Salary	None	None	8	Right	Scale

Notice that the type specified for Annual Salary is numeric. This does not change how the data can be used but if you prefer to view these data with dollar symbols, you can change the type specified for this variable to "Dollar."

Now you are ready to enter your data. Click on the Data View tab toward the bottom of the spreadsheet. This tab of the spreadsheet has several columns into which you can enter the data for each variable. Each column will show the names given to the variables that were entered previously using the Variable View tab. Click the first cell corresponding to the first row of the first column. Start here and follow these steps:

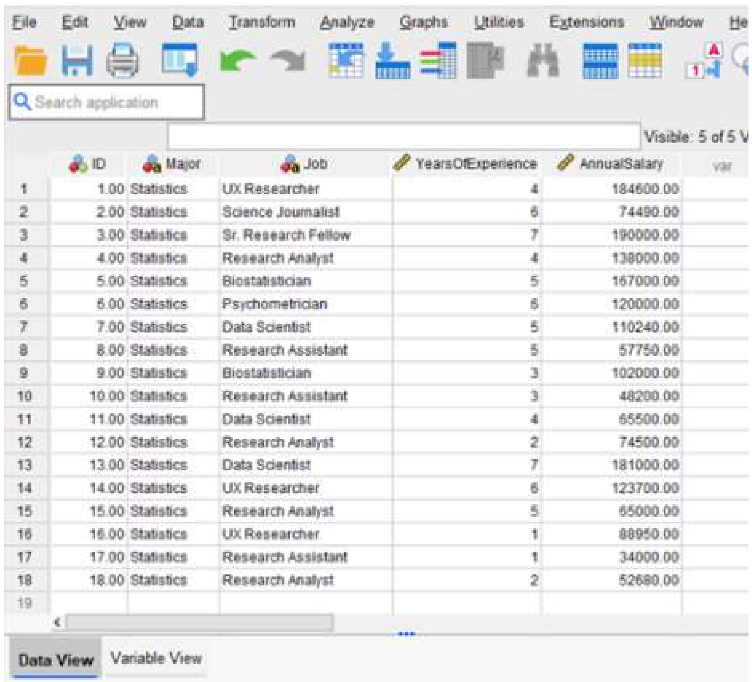
1. Enter the data for the first variable moving down the rows under the first column. If your data are already on your computer in a spreadsheet format such as excel, you can copy-paste the data in for the variable. Note: If the spreadsheet will not allow you to enter the information and/or makes a blunt tone when you try to enter the data, it means you have an error in the format of the

data you are trying to enter or that they do not match the details you provided in the Variable View tab. Change the data format or go back to edit the information on the Variable View tab if this occurs, as appropriate. Then, return to the Data View tab to enter your data.

2. Repeat step 1 for each variable until all of your data have been entered.
3. Then hit save to ensure your data set will be available for you in the future.

> **Note**
>
> Data entered into SPSS are saved in a file format that can only be opened in specific forms of software such as SPSS. Therefore, if you use a computer at school or work that has SPSS to make data files and try to open them on a different computer which does not have SPSS, you will not be able to. Keep this in mind if you plan to use different devices while practicing the use of SPSS and statistics.

Here is what the Data View tab would look like when created for Data Set 2.1:



	ID	Major	Job	YearsOfExperience	AnnualSalary	var
1	1.00	Statistics	UX Researcher	4	184600.00	
2	2.00	Statistics	Science Journalist	6	74490.00	
3	3.00	Statistics	Sr. Research Fellow	7	190000.00	
4	4.00	Statistics	Research Analyst	4	138000.00	
5	5.00	Statistics	Biostatistician	5	167000.00	
6	6.00	Statistics	Psychometrician	6	120000.00	
7	7.00	Statistics	Data Scientist	5	110240.00	
8	8.00	Statistics	Research Assistant	5	57750.00	
9	9.00	Statistics	Biostatistician	3	102000.00	
10	10.00	Statistics	Research Assistant	3	48200.00	
11	11.00	Statistics	Data Scientist	4	65500.00	
12	12.00	Statistics	Research Analyst	2	74500.00	
13	13.00	Statistics	Data Scientist	7	181000.00	
14	14.00	Statistics	UX Researcher	6	123700.00	
15	15.00	Statistics	Research Analyst	5	65000.00	
16	16.00	Statistics	UX Researcher	1	88950.00	
17	17.00	Statistics	Research Assistant	1	34000.00	
18	18.00	Statistics	Research Analyst	2	52680.00	
19						

Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

Generating Histograms

The steps to generating a histogram in SPSS are:

1. Click Graphs -> Histogram.
   a. In some versions of SPSS the histogram option is a bit more hidden. If you are using a version where it does not appear in the initial graphs option menu, use these steps to generate a histogram in SPSS: Click Graphs -> Legacy Dialogs -> Histogram.
2. Drag the name of the variable you want to plot as a histogram from the variable list on the left into the Variable text box on the right of the command window. You can also do this by clicking on the variable name to highlight it and the clicking the arrow to move the variable from the left into the Variable text box on the right.
3. Click OK.
4. The histogram will appear in the SPSS output viewer but may not look the way you want it to.

To edit your histogram, follow these additional steps:

5. Double-click the histogram to open the chart editor.
6. Double-click on any anchor under the x-axis to begin editing the presentation of scores or intervals along this axis.

a. Click “Scale” if you want to adjust the range shown and the size of intervals. Enter the values you wish to use in the “Custom” text boxes. Click on “Number Format” if you want to change how many decimal places are shown for values along the x-axis. Fill in the text boxes to indicate the desired number of decimals to show. Click “apply” to see the graph adjusted to your specifications. Repeat these steps until you are satisfied with the organization of the x-axis.

7. Double-click on any number along the y-axis to begin editing the specificity of frequencies. a) “Click “Scale” if you want to adjust the specificity of frequencies shown along the y-axis. Enter the values you wish to use in the “Custom” text boxes. Click on “Number Format” if you want to change how many decimal places are shown for values along the y-axis. This is rarely needed as the default is to show to the whole number which is appropriate for most data. When whole numbers are not desired, fill in the text boxes to indicate the desired number of decimals to show. Click “apply” to see the graph adjusted to your specifications. Repeat these steps until you are satisfied with the organization of the y axis.
8. Double-click on any bar of the histogram to begin editing the appearance of the bars. a) Click “Fill & Border” if you would like to change the colors of the bars or their borders. Select the desired colors. You can also change the style and width of the borders here. To change the thickness of the borders, use the pulldown window under “Weight.” To Change the style of the bar borders, use the pulldown window under “Style.” Click “apply” to see the graph adjusted to your specifications. Repeat these steps until you are satisfied with the appearance of the bars of the histogram.

Generating Simple Bar Graphs

The steps to generating a univariate bar graph in SPSS are:

1. Click Graphs -> Bar Graph -> Simple -> Define
2. Drag the name of the variable you want to plot as a bar graph from the left into the Category Axis text box on the right of the command window. You can also do this by clicking on the variable name to highlight it and the clicking the arrow to move the variable from the left into the Category Axis text box on the right.
3. Click OK.
4. The bar graph will appear in the SPSS output viewer but may not look the way you want it to.

To edit your bar graph, follow these additional steps:

5. Double-click the bar graph to open the chart editor.
6. Double-click on any anchor under the x-axis to begin editing the presentation of categories shown along this axis.
    a. Click “Categories” if you want to adjust the order of the categories shown. Each category will be listed under “Order.” Drag the names of the categories into their desired order. You may also want to change the angle at which the category names appear under the x-axis. To change this, click on “Labels & Ticks.” Then select the preferred orientation from the pull-down menu under “Label orientation.” This is where you can choose whether you want the names to appear vertically, horizontally, or at an angle under the x-axis. Click “apply” to see the graph adjusted to your specifications. Repeat these steps until you are satisfied with the organization of the x-axis.
7. Double-click on any number along the y-axis to begin editing the specificity of frequencies.
    a. Click “Scale” if you want to adjust the specificity of frequencies shown along the y-axis. Enter the values you wish to use in the “Custom” text boxes. Click on “Number Format” if you want to change how many decimal places are shown for values along the y-axis. This is rarely needed as the default is to show to the whole number which is appropriate for most data. When whole numbers are not desired, fill in the text boxes to indicate the desired number of decimals to show. Click “apply” to see the graph adjusted to your specifications. Repeat these steps until you are satisfied with the organization of the y axis.
    b. If you would like the y-axis to be labeled as “frequency” instead of “count.” Double-click on the word “count” where it appears beside the y-axis. This will activate the textbox for the y-axis label so that you can edit it. Click outside of this label textbox when you are satisfied with the name you have given to the y-axis.
8. Double-click on any bar of the histogram to begin editing the appearance of the bars.
    a. Click “Fill & Border” if you would like to change the colors of the bars or their borders. Select the desired colors. You can also change the style and width of the borders here. To change the thickness of the borders, use the pulldown window under “Weight.” To Change the style of the bar borders, use the pulldown window under “Style.” If you want to change the width of the gaps between bars, click “Bar Options.” Use the slide ruler under “Bars” to indicate the desired gap width. Click “apply” to see the graph adjusted to your specifications. Repeat these steps until you are satisfied with the appearance of the bars of the histogram.

2.9: Chapter References

Wu, H., & Leung, S. O. (2017). Can Likert scales be treated as interval scales?—A simulation study. *Journal of Social Service Research*, *43*(4), 527–532.

https://doi.org/10.1080/01488376.2017.1329775

CHAPTER OVERVIEW

3: Measures of Central Tendency

One group of quantitative techniques that are used to summarize data are known as *Measures of Central Tendency.* Another term for measures of central tendency is *averages.* As the name implies**, measures of central tendency** summarize where a group of scores (i.e. data for a variable) tended to fall. The goal of using a measure of central tendency is to simplify by representing all the data for a variable with just one or a few numbers or terms. A separate, but related, set of techniques known as *Measures of Variability* summarize the extent to which a group of scores tended to be different from each other. The foci of this chapter are the three commonly used measures of central tendency and an introduction to deviation, a concept that underlies some measures of variability which we will learn more about in Chapter 4.

Once we have our raw scores, the statistical work must begin. A single piece of data is easy to understand. For example if someone said his age was 18, the datum for age is simply 18. This can be summarized with the statement, "His age is 18." This statement, which summarizes one datum, is simple. However, statisticians and researchers often need to collect lots of data to understand a variable more broadly. This means that several cases are often measured, each of which will have a raw score for the variable. If we are using these data to understand the variable, then the data will typically be summarized in some way. Summarizing the data is preferable to considering each datum individually because the goal is usually to understand the variable.

3.1: The Goal of Summaries

When summaries are created, a statistician must balance two competing goals: gaining simplicity and retaining specificity. To simplify data means to reduce what is shown or stated to emphasize a key aspect or aspects of the data or the variable the data represent. To specify data means to retain and show details about the data or the variable the data represent. These two goals are in conflict with each other and some of the work a statistician must do is aimed at balancing the two so that simplicity is gained without losing too much specificity. Different summaries can be created, some of which emphasize simplicity over specificity and others that focus on adding some specificity back in. The measures of central tendency are focused primarily on simplicity rather than specificity. However, there are different things that one can focus on when summarizing data for simplicity. Each measure of central tendency focuses on one aspect of the data when summarizing. Therefore, we must understand the focus of each measure of central tendency to know which one will provide the best summary for a given situation.

Take a look at Data Set 3.1; there are 22 scores for age. It is hard to state anything succinct about age when we have 22 pieces of information about age in their raw form. The more data we have, the harder it gets to simply eyeball what we have in the dataset. Imagine how hard it would be to look at data for age from 10,000 cases and get a clear sense of what the data showed. To reduce the strain, data can be summarized with a measure of central tendency rather than simply viewed in their raw form. Take a moment to look at Data Set 3.1 and try to describe the data for age succinctly.

Data Set 3.1

Age
47
46
42
39
36
34
33
33
32
29
29
29
28
27
25
23
20
19
19
18
16
14

How did you describe or summarize the data for age in Data Set 3.1? You may have noticed that age varied and included teens through adults. You may have noticed that the lowest age was 14 and that the highest age was 47, that most people were in their teens, 20s or 30s, and that only 3 people were in their 40s. You may have even noticed that most ages were only reported once but that 29 was reported three times. Though these statements are true, they all differ and none of them provides a single number or description that can be used to summarize all of the data. In addition, different individuals might focus on a different datum when summarizing and may use different words to describe what they notice in the data. These differences can create confusion when several individuals are trying to understand the same variable, even when they are looking at the same data for that variable. These differences can make it especially difficult to understand the similarities or dissimilarities across different datasets. Therefore, it is helpful to have tools, such as measures of central tendency that allow us to make simple descriptive statements.

Measures of central tendency are strategies used to summarize raw data. Though there are many, we will focus on three common measures of central tendency: the mode, the median, and the mean. Each of these is calculated or identified in a systematic way. One of the benefits of using one of these techniques is that others will recognize the name of the measure of central tendency and know exactly which calculations were used to summarize the data. This helps improve clarity when summaries are shared with others, including large audiences.

3.2: Mode

The **mode** refers to the number or response which occurs most frequently. The mode can be used to summarize quantitative or qualitative data. This makes it more flexible than the median and the mean. In this section the terms “numbers,” “scores,” or “quantitative responses” will be used to refer to quantitative data and the terms “qualitative responses” or “categories” will be used to refer to qualitative data.

There are three important rules to finding the mode:

1. The data must be varied rather than constant.
2. The mode must be the number or qualitative response that occurred the most frequently.
3. The mode must be a number or qualitative response that occurred more than once.

When these three criteria are not met, there is no discernable mode. In this case we can report the results as follows: “There was no mode.” If the first rule is not met, a mode should not be reported. Instead, it should be stated that the data were constant. If the third rule is not met, providing a mode would be useless. This is because all raw score are the mode when rule three is not met and, thus, we would simply be restating the full data set as the mode. When this occurs no simplicity or reduction of information is being gained through reporting the mode.

Sometimes the second criteria is met by two different scores or responses. When this occurs both are considered the mode. If the second criteria is met by three different scores or responses, then all three are considered the mode. Thus, it is possible to have multiple modes in which case all are reported together to as the mode when summarizing a data set.

Modality

If you have data in which two numbers or qualitative responses occur more than once and equally often, your dataset has two modes and is referred to specifically as *bimodal*. If two or more such numbers or qualitative responses occur, you can refer to the data more generally as *multimodal*. Data are often expected to be unimodal for reasons which will be covered in Chapter 5. For now, our focus is on identifying modes and modalities.

Modality

Modality can be defined as the way the data conform to a mode. In other words, the modality is how the mode is seen or experienced. There are different modalities that can be observed in a dataset. Sometimes there is only one mode for a dataset, other times there may be two, three, four, or even one hundred modes! Specific terms are used to refer to the modality of a dataset.

- *Unimodal* means there is only one mode.
- *Bimodal* means there are two modes.
- *Trimodal* means there are three modes.
- *Multimodal* is a broad term that means that there is more than one mode.

Multimodal is often used whenever there are more than three modes but it can also be used in place of bimodal or trimodal.

The Mode for Quantitative Data

The mode is found for quantitative data using the three important rules. Look back to Data Set 3.1 and use the three rules to try to deduce whether there is a mode and, if so, what the mode is.

1. The data were not all the same.
2. The number that appeared the most often was 29.
3. This number occurred more than once.

The mode for age in Data Set 3.1, therefore, is 29. This is considered unimodal meaning there is one (uni) mode. The symbol often used for the mode is *Mo.* The same symbol is used when reporting the mode for a sample as a mode for a population. In addition, APA-format dictates that values should be rounded and reported to the hundredths place in most circumstances. Therefore, the mode for age should be shown to the hundredths place. Therefore, the result for the mode can be written using symbols as:

$$\boldsymbol{Mo} = \mathbf{29.00}$$

The Mode for Qualitative Data

The mode is a bit special because it can be used with many different kinds of data, including qualitative data which are on the nominal scale of measurement. Recall that nominal data refer to responses that represent categories or characteristics rather than amounts. For example, we may want to know what various students chose as their majors. Major is not represented with numbers; instead, it is represented with a word or group of words (such as psychology, nursing, mathematics, biology, child development, or media production). These kinds of data are referred to as *nominal* which means "name." This makes sense if you think about the fact that when people are asked their majors, which is nominal, they respond by providing the name of their majors.

Data Set 3.2

Gender
Female
Male
Male
Female
Transgender
Transgender
Male
Male

The mode is great for summarizing nominal data. Take a look at Data Set 3.2. These data represent the genders of 8 individuals. Gender data are nominal data because they cannot be meaningfully represented in numbers. Notice that words are used to represent the data for gender. Use the three rules for finding the mode to answer the question: What is the *mode* for Gender in Data Set 3.2?

1. The data were not all the same.
2. The qualitative response that appeared the most often was *Male*.
3. This qualitative response occurred more than once.

Thus, the mode is Male. This is considered *unimodal* because there is only one mode for gender. We can also write this using the symbol for mode as:

$$\boldsymbol{Mo} = \textbf{Male}$$

Reading Review 3.1

1. What are the two competing goals statisticians consider when creating summaries?
2. What are the three rules to finding a mode?
3. Find the sample size, mode, and modality for each dataset:

Daily Hours of T.V.			
8.00	4.25	0.00	5.75
3.50	1.00	6.25	7.00
1.00	5.50	4.00	2.00
2.50	0.00	3.50	4.00
0.00	4.00	1.50	0.50
5.00	2.50	2.75	10.00
4.00	3.50	0.00	1.00
0.00	1.25	4.00	2.00

Test Scores	
96	82

Test Scores	
94	80
91	79
90	73
89	71
88	65
85	52
84	47

Favorite Dessert
Oatmeal Cookies
Fruit
Chocolate Cake
Carrot Cake
Chocolate Cake
Ice Cream
Apple Pie
Chocolate Cake
Brownies

3.3: Median

The **median** is a measure of central tendency that refers to the score that is in the middle of an ordered set of quantitative scores. Essentially, a median is a midpoint of ordered scores. This means that about half of the scores represent values greater than the median and the other half represent values lesser than the median. Therefore, think of the median as representing the dividing line between the upper and lower 50% of scores rather than representing a score that was common, existed in the data, or even could have existed in the data. The median can only be used with data for quantitative variables; qualitative data cannot have a median because qualitative responses cannot be organized in order of magnitude. Just like is true for the mode, a median should only be reported for variables and not for constants. In addition, if the variable is quantitative but discontinuous, medians should be used and interpreted with caution.

When there is an odd number of scores, the number in the middle of the set of scores is the median. To find the median for a data set with an odd sample size, first ensure that you have data for a variable rather than a constant. Then, follow these two steps:

1. Organize the raw scores from smallest value to highest value.
2. Find the number that falls exactly in the middle of those ordered scores. This number is the median.

Where there is an even number of scores, two numbers are in the middle so an extra step is needed to find the median. In these cases, the median is the midpoint *between* the two scores that are on either side of the middle of the ordered scores. To find the median for a data set with an even sample size, first ensure that you have data for a variable rather than a constant. Then, follow these three steps:

1. Organize the raw scores from smallest value to highest value.
2. Find the two numbers that fall on each side of the middle of those ordered scores.
3. Add these two numbers and divide the sum by 2. The resulting number is the median.

The third step allows us to find the midpoint between two scores when that midpoint was not observed in the data.

Data Set 3.3 (Miles Per Hour)

Raw Data	Reordered Data
65	81
77	77
68	77
76	76
81	75
64	72
65	68
77	65
72	65
75	64

Let’s take a look at an example using Data Set 3.3. These data represent speeds driven on the freeway in miles per hour. This is quantitative and measured on a ratio scale. The first column shows the data before being organized. The first step is to put the data in order from lowest to highest which is shown in the second column for Data Set 3.3. The sample size is 10. A dashed line is shown in the ordered data column to indicate where the median score would be in the list of scores. Notice that the midpoint falls between two of the scores. Because the sample size is even, we must find the midpoint between the two scores which are closes to the middle. These scores are 75 and 72. Therefore, the midpoint between 75 and 72 is the median and can be computed by summing these values and dividing by 2 as shown:

Median Calculations Using Data Set 3.3

$$\mathrm{Mdn} = \frac{75+72}{2} = \frac{147}{2} = 73.50$$

There are a variety of symbols and abbreviations that are used for the median which include, for examples, *Mdn*, *Med,* and $\tilde{x}$. The symbol $\tilde{x}$ is sometimes called x-tilde because it has a tilde over the top of it. Any of these can be used but for this chapter we will use *Mdn* for simplicity. The same symbol is used when reporting the median for a sample as a median for a population.

Mdn = 73.50

Look back at Data Set 3.1 for age. There are 22 scores for the variable age. This is an even number of scores which means there is no number exactly in the middle; in a case like this, you find the point between the two middle scores. The two scores in the middle of the list (11th from the top and 11th from the bottom) are 29 and 29. This makes things easy because the midpoint between 29 and 29 is just 29! Following APA guidelines, median is generally reported to the hundredths place; thus, our median would be reported as 29.00. Recall from earlier that the mode for the data in Data Set 3.1 was also 29.00. Thus, we have used two measures of central tendency that have yielded the same summary number for age in dataset 3.1.

Median Calculations Using Data Set 3.1

$$\mathrm{Mdn} = \frac{29+29}{2} = \frac{58}{2} = 29.00$$

3.4: Mean

The mean is often what people are referring to when they say "average," though this is not the only average that can be used (as we have seen, the mode and median are also averages). The **mean** is found by adding all the scores in a dataset and dividing that total by the number of cases in the dataset. The mean is balanced such that the total of how far people were below the mean is equal to the total of how far people were above the mean. This is a very important feature of the mean.

Statistical Notation

Statistical notation refers to the various symbols which are used to represent concepts in formulas and when reporting results. The formulas for calculating the mean are shown below. One formula is for a sample and the other is for a population. The calculations are the same but we use different symbols to clarify whether our data were drawn from a sample (i.e. a subset, or part, of a population) or the population (i.e. all cases). When using a formula to find a sample mean, $\bar{x}$ is used. This symbol is often referred to as "X bar." However, *M* can also be used as the symbol for the mean. When using a formula to find a population mean, μ is used; this is the lowercase Greek letter whose name is mu. The other difference between the formulas is that we use a lowercase *n* for a sample to show that it is not the whole group (i.e. it is smaller than could be possible if every case was included) and capital *N* for the population to show that it includes all cases. Though the symbols are different, the steps of the sample and population mean formulas are the same.

Formulas for the Mean

Sample Mean	Population Mean
$\bar{x} = \frac{\Sigma x}{n}$	$\mu = \frac{\Sigma x}{N}$

Calculating the Mean

The steps for calculating a mean are as follows:

1. Add up all of the raw scores to solve for Σx
2. Divide Σx by the sample size.

That is it! There are only two steps and they must be done in the order specified.

Try using the formula for the mean with the age dataset from Data Set 3.1. Follow the two steps:

1. **Add up all the raw scores**

$$\Sigma x = 14+16+18+19+19+20+23+25+27+28+29+29+29+32+33+33+34+36+39+42 \\ +46+47=636$$

$$\Sigma x = 638$$

2. **Divide Σx by the sample size**

$$\sum \boldsymbol{x} = 638 \text{ and } n = 22$$

$$\bar{x} = 638/22 = 29.00$$

M = 29.00

Let's practice the formula again using Data Set 3.3. This time we will show all the data in the formula at once rather than broken out into steps:

<u>Mean Calculations Using Data Set 3.3</u>

$$\bar{x} = \frac{64+65+65+68+72+75+76+77+77+81}{10} = \frac{720}{10} = 72.00$$

When we calculated the means for Data Set 3.1 and Data Set 3.3, the results were whole numbers (though we should still report them to the hundredths place). However, this will not always be the case. Let's demonstrate this by finding the mean for Data Set 3.4.

<u>Mean Calculations Using Data Set 3.4</u>

Data Set 3.4

Height in Inches
74
72
71
68
66
65
65
64
63
62
61
59

$$\bar{x} = \frac{59+61+62+63+64+65+65+66+68+71+72+74}{12} = \frac{790}{12} = 65.833\bar{3}$$

Rounding and Symbol Use Guidelines

The mean for height in Data Set 3.4 has a repeating decimal (also known as a recurring decimal). Rational numbers with repeating decimals (like the mean for height) and irrational numbers continue infinitely. Generally, however, infinite specificity is beyond what is needed when reporting results. Without conventions, statisticians could choose to round and show this mean a variety of ways which would make comparisons of data from different

reports harder to compare. Some fields, therefore, have developed rules and guidelines for the place to which values should be rounded and shown. These guidelines help ensure consistency and comparability of results across studies. Psychology, for example, follows the guidelines set forth in the American Psychological Association (APA) style manual. APA guidelines state that most values found through a process or formula (which are often referred to as *obtained values*) should be rounded and reported to the hundredths place. An exception is made when data were measured using an interval scale, for which, rounding to the tens place is acceptable under some conditions (APA, 2022). This meets the two objectives of summarizing by providing some specificity (two decimal places) while gaining simplicity (by not having a number continue to trail infinitely).

Further, various symbols are used for the mean depending upon the situation. When reporting means in sentences or tables, APA guidelines state that the symbol *M* be used in place of $\bar{x}$ or μ. M is the uppercase version of the lowercase Greek letter μ and is also the first letter of the word “mean.” *M* is much easier to use when typing because it is on a Standard English keyboard and does not require inserting a special symbol into a document. For these reasons, it was a logical symbol to use for abbreviating the word “mean.”

If the APA guidelines are followed, the obtained value for Data Set 3.4 could be written as such:

***M* = 65.83**

This guideline for rounding only applies to the final result, not the raw data or steps of the formula. Keep in mind that when we round numbers we are losing some specificity and can be introducing a little bit of error in the process. Thus, the fewer decimal places you show when rounding, the greater the potential rounding error can be. The sooner this error is introduced in the computations, the greater its potential to impact the final result. Therefore, though final results are often rounded to the hundredths place, if a value is being used as a step in another calculation or formula it is best to use the exact, unrounded value or to keep steps to at least four decimal places if rounding.

Comparing the Mean to the Mode and Median

The mean, median, and mode are three options for finding and reporting central tendencies. Though all three can be used to summarize data for quantitative variables, only one is typically used at a time. There are two main reasons for this. First, reporting all three goes against one of the goals of summarizing which is to simplify. Second, the mean, median, and mode are often expected to yield similar results, for reasons we will review in detail in Chapter 5. Therefore, reporting all three measures of central tendency will yield redundant summaries in some cases. Third, when the mean, median, and mode do differ, it can be due to a relevant feature of the variable which is better captured or dealt with using one of the three measures of central tendency. Thus, it is necessary to understand the differences which underlie the three measures to know which one is the best fit for each data set or variable.

Balancing Deviations

The mean is the point at which deviations are balanced. **Deviation** refers to the difference between the mean and any of the raw scores**.** The formula for the mean produces a number that is above some raw scores and below others. This means that some raw scores will have a positive deviation because they are greater in value than the mean while others will have a negative deviation because they are lesser in value than the mean. The deviation balances positive and negative deviations such that if all the deviations are added the sum will be 0. This is what is meant when we say the mean balances deviations.

Look back at Data Set 3.1 keeping in mind that the mean was 29.00. Notice that several of the raw scores for age such as 16 and 20 are below the mean, and others such as 32 and 44 are above the mean. If you find the deviations for all of the raw scores below the mean and then add them up, you will get the same absolute value as when you find all the deviations for the raw scores above the mean and add them up. Therefore, the mean is the point (or value) that balances between the total deviations below it and the total deviations above it.

Let's look at the balance point property of the mean using our data for age (Data Set 3.1). We can demonstrate the way the mean functions as a balance point by subtracting the mean from each raw score to find the deviations (see Table 1). The first column of Table 1 lists all of the raw scores. The second column lists the mean for the dataset. The third column lists the deviation for each raw score. When you sum the deviations, the result is 0. This is because the deviations below the mean balance with the deviations above the mean and will always sum to 0. This is what is being stated when we refer to the mean as a balance point; the mean is a number that causes the sum of the deviations to balance out to 0.

Formula for Deviation

The deviation for an individual score is calculate as:

$\text{dev} = x - \bar{x}$

x refers to an individual score and $\bar{x}$ refers to the mean.

The total deviation for a dataset is calculated as:

$\text{ Sum of dev } = \sum(x - \bar{x})$

$\sum$ indicates that all deviations should be added.

Table 1 Deviations for Data Set 3.1

Raw Score	Mean	Deviation*
47	29	18
46	29	17
42	29	13
39	29	10
36	29	7
34	29	5
33	29	4
33	29	4
32	29	3
29	29	0
29	29	0
29	29	0

Raw Score	Mean	Deviation*
28	29	-1
27	29	-2
25	29	-4
23	29	-6
20	29	-9
19	29	-10
19	29	-10
18	29	-11
16	29	-13
14	29	-15
		Sum of Deviations = 0

*Note

Raw scores are shown to the whole number to reflect how they appeared in the data set and the steps are also shown to the whole number because all decimal places were zeros.

The mean and median are both balance points but focus on balancing two different things. The mean balances deviations, whereas the median balances the number of cases. Take a look at the data. We see that 12 scores were above the mean and 10 scores were below the mean. Thus, unlike the median, the mean doesn't always balance the number of cases that are above or below the mean; instead, the mean refers to how far those scores are from the mean. In contrast there is a balance of cases from the median, with 11 cases above the median and 11 cases below the median.

Note that the median point represents the point between the bottom two scores of 29 causing one of the 29s to represent a value below 29 and three to represent values above 29. This is because the real score limits (or boundaries of what rounds to 29) are from 28.50 to 29.50 and, thus, the 29s are considered to represent various places in this range. Thus, sometimes a different version of the median known as the interpolated median (also known as the precise median) are used to better represent this. However, the standard median is used more often in the behavioral sciences and, thus, is the focus of this chapter. Those interested in the way a median serves as a balance of cases when the same value appears on each side of the middle value (like we see in Data Set 3.1) are encouraged to review the interpolated median online or, if you have a professor like Dr. Peter who loves explaining how numbers are conceptualized when using medians, ask your professor during their office hours. If they have read and recommended this book to you, they should see this question coming.

The Impact of Outliers and Sample Size on the Mean. Because the mean balances deviations, it is sensitive to lack of symmetry in data or something known as an outlier. Outliers are rare (meaning infrequently occurring) scores that are more extreme than the other scores in a dataset. When an outlier is present, the mean has to shift towards the outlier to balance the deviations. This shift is called *skew*. Let's compare two data sets to see an example of how an outlier shifts data. Data Set 3.5 includes data for the variable Household Income from a small sample of 10 cases. The incomes range from $47,000 to $112,000. Though the range of incomes is fairly

large, none of the incomes reported are extremely different than the next closest score or the data set overall. The mean of income for Data Set 3.5 is $80,000. The scores vary with some incomes being close to the mean and others farther from the mean.

Data Set 3.5 Household Income in Dollars (n = 10)

Raw Score	Mean	Deviation
112,000	80,000	32,000
103,000	80,000	23,000

91,000	80,000	11,000
87,000	80,000	7,000
80,000	80,000	0
80,000	80,000	0
75,000	80,000	-5,000
66,000	80,000	-14,000
59,000	80,000	-21,000
47,000	80,000	-33,000
		Sum of Deviations = 0

Now let's look at Data Set 3.6. Data Set 3.6 includes all 10 scores for the variable Household Income as Data Set 3.5 but with the addition of an 11th case with an income of $1,950,000. The score of $1,950,000 is an outlier. Let's take a look at what happens to the statistics for the data when this outlier is present. The incomes in Data Set 3.6 range from $31,000 to $1,950,000. The mean of income for Data Set 3.6 is $250,000, which is much higher than the mean was when the outlier was not in the data (as shown in Data Set 3.5). This is because the mean had to move closer to the outlier to allow the deviations to balance out to zero. The deviations for all raw scores in data set 3.6 are also very large. This is because the outlier has dramatically impacted the mean.

Data Set 3.6 Household Income in Dollars (n = 10)

Raw Score	Mean	Deviation
1,950,000	250,000	1,700,000
112,000	250,000	-138,000
103,000	250,000	-147,000
91,000	250,000	-159,000
87,000	250,000	-163,000
80,000	250,000	-170,000
80,000	250,000	-170,000
75,000	250,000	-175,000
66,000	250,000	-184,000
59,000	250,000	-191,000
47,000	250,000	-203,000
		Sum of Deviations = 0

The name for this kind of impact is skew. **Skew** refers to the asymmetry in quantitative data which can have a greater impact on the mean than the median or the mode. The mean can be greatly impacted by an outlier because outliers pull the mean toward them in order to balance deviations. The measures of central tendency for data sets 3.5 and 3.6 are summarized in Table 2. Compare the mean for the sample without an outlier (Data Set 3.5) which was $80,000 to the mean for the sample with an outlier (Data Set 3.6) which was $250,000. The outlier is a very high score and when it was added to the data set, it caused the mean to increase. This is known as positive skew. **Positive skew** refers to an increase in the mean due to asymmetry in data caused by a high score or high scores. The mean is pulled up toward the high outlier. Consistent with this, a low outlier will cause negative skew. **Negative skew** refers to a decrease in the mean due to asymmetry in data caused by a low score or low scores.

The median, by comparison, is less likely to be dramatically impacted by outliers and asymmetry because it balances cases rather than deviations. This allows the median to experience more stability when an outlier is added to a data set. To illustrate this, let's compare the medians for the data when with (Data Set 3.6) and without the outlier present (Data Set 3.5). The medians are the same for Data Set 3.5 and Data Set 3.6 (see Table 2). Thus, we can see that, unlike the mean, the median is not very sensitive to the impact of an outlier.

The presence of skew can be identified by comparing means to medians. When the mean and median are equal, it indicates that the mean has not been skewed. When the mean is higher than the median, it indicates that there is positive skew. This is because the mean gets pulled in the direction of an outlier more so than the median is pulled. For this same reason, when a mean is lower than a median, it indicates that there is negative skew in the data set.

Table 2 The Impact of an Outlier on Measures of Central Tendency

	Without Outlier	With Outlier
Mode	80,000	80,000
Median	80,000	82,000
Mean	80,000	250,000

The Role of Sample Size

Sample size can mitigate or aggravate the impact of an outlier. The impact of an outlier is based on two things:

1. How extreme the outlier is relative to the rest of the scores and
2. Sample size.

The more extreme the outlier is, the greater its impact can be. However, the impact of an outlier is proportional to how much of the sample is comprised of the outlier. When the sample size is small, there are fewer values to balance against an outlier which allows the outlier to have a greater impact. You can see this demonstrated in Data Set 3.6 which includes an extreme outlier in a small data set. The only score above the mean is the outlier and all other scores had to balance it out by being below the mean. When the sample size is large, there are more values to balance against the outlier which mitigates its impact. Thus, the outlier in Data Set 3.6 would have more ability to skew the mean in a sample of 11 cases where it represents one eleventh of the observations than in a sample of 100 cases where it represents only one-hundredth of the observations. You can summarize it this way: When the sample size is small, the outlier represents a larger proportion of the scores and, thus, has more power to move the mean; however, when the sample size is large, the outlier represents a smaller proportion of he scores and, thus, has less power to move the mean.

Reading Review 3.2

1. What is the sample size for the test score data?
2. What is the median for the test score data?
3. What is the mean for the test score data?
4. Does there appear to be any positive or negative skew in the test score data?

Test Scores	
96	82
94	80
91	79
90	73
89	71
88	65
85	52
84	47

Choosing between the Mode, Median, and Mean. There are several things a statistician should consider when deciding which measure of central tendency to use to summarize a variable. Here, we will review a few of the main considerations.

A statistician must ensure that the measure of central tendency is appropriate for the way a variable was measured. If data are qualitative, the mode is the only option. The mode is also often considered appropriate for ordinal data because of quantitative limitations with this scale of measurement. The mode is particularly useful when a score or qualitative response repeats several times in a data set. If there is no mode, it is generally best to choose the median or the mean for interval- or ratio-level data unless the goal of the summary is to indicate that there were no repeating scores or qualitative responses.

If the variable is measured on the interval or ratio scale, the mode, median, or mean could all be appropriate so additional factors should be considered. The statistician should consider which measure of central tendency will provide the most appropriate information for their current goals. Thus, the mean is most commonly chosen for quantitative data because it can be used to serve many different goals. This is because the mean is a foundational descriptive statistic on which more advanced, inferential statistical procedures such as t-tests (which are covered in Chapters 7 through 9) and ANOVAs (which are covered in Chapters 10 and 11) rely. However, when there is problematic skew, the median is generally recommended and used instead of the mean.

Measures of Central Tendency Compared to Real Scores. The mode has one benefit that the median and mean do not. When there is a mode, it will always be the same as at least two observed scores or qualitative responses. The mode for a sample cannot be something that was not observed in the sample data. Keep in mind that measures of central tendency are meant to summarize the data. Another way of thinking about a summary is that it is meant to describe what was generally true. However, only the mode is guaranteed to provide a summary that was true for at least two cases.

In contrast, the median and the mean for sample data can be values that were not true of any observed cases. Take the median for Data Set 3.3 which was 73.50. We can summarize those data by saying the median speed was 73.50. However, there isn’t a single case where the speed was exactly 73.50 miles per hour. Take a look at

the mean for Data Set 3.5. Therefore, it is important to keep in mind what each summary is and is not able to tell us. The measures of central tendency are not designed to tell what is always true. They can’t because they are built for summarizing things that vary not things which are constant. The median and mean aren’t even designed to tell us what was at least sometimes true (though they do sometimes yield values that occurred in the data set). Instead, they are each designed to focus on a different way of stating what was approximately (or tended to be) true. This can lead to some confusion when we read statements such as “The average household has 2.50 children” because no household can actually have 2.50 children as you cannot have a fraction of a person. What is meant when an “average” is stated in this way is a theoretical and approximate representation of what tended to be true. Note also that the word “average” in the sentence is vague as the mode, median, and mean are all averages. We know that the average being referred to is not the mode because the mode can only be a value or qualitative response which can and did occur. That means that the average being used to summarize number of children is either the median or the mean, but we cannot be sure which. This is why it is always important to specify when reporting results.

3.5: Reporting Results

Results in the behavioral and social sciences are often reported in sentences or, for more complex analyses, paragraphs. When statisticians report their results, they must take care to do so accurately and clearly. To be accurate means that what is reported is correct and that there have been no miscalculations. This requires that the statistician takes care to learn the proper procedures and formulas and that they carry out the steps with precision. It is also a good habit to double-check one's work again before reporting the results, the same way it is a good idea to proofread a paper before submitting it to one's professor. Clarity refers to reporting things in a way which is comprehensible. In practice, this means that results sentences should be specific rather than vague and that they be written simply and directly whenever possible. As noted earlier in this chapter, APA format provides some useful guidelines which are meant to enhance clarity such as the rule that decimals be rounded and shown to the hundredths place when reported as results. When reporting results in sentences, do each of the following:

1. Ensure results are accurate and shown to the proper decimal place.
2. Replace any vague words or language with specific terms and phrases.
3. Edit for simplicity where possible such as by removing unnecessary words.

Understanding statistics, research, and sciences can sometimes be challenging so it is beneficial to audiences when writing is direct.

Let's take a look at an example of a results sentence that would benefit from revision.

When the average was computed for the data using the mean formula, the result for the variable age was 29.0000000.

Reading this results sentence is painful, but it doesn't have to be this way. Let's edit following the three recommended steps from above. First, let's round following APA format:

When the average was computed for the data using the mean formula, the result for the variable age was 29.00.

It's starting to look better but could still be improved. Let's identify any vague language. The word "average" is not specific so it should be replaced. The measure of central tendency which was used was the mean so let's make that clear:

When the mean was computed for the data using the mean formula, the result for the variable age was 29.00.

We are getting there but this sentence is cumbersome. We want to focus on the main points so let's identify what those are. We want to indicate what the variable was. It was age. We want to indicate which measure of central tendency was used. It was the mean. Finally, we want to state what the statistic was. It was 29.00. The rest of the information in the sentences is unnecessary. We don't need to state that a computation occurred because our audience will know this simply because we are reporting the result of that computation. We also don't need to say we are giving a result; instead we can simply give the result. Further, we do not need to state that the data were for a variable because means are only computed for variables. Taken together, this means we can remove several parts of the sentence to leave just the things we need like so:

When the mean was computed for the data using the mean formula, the result for the variable age was 29.00.

Now we have a simple and clear results sentence which states:

The mean age was 29.00.

Much better. Reread the initial sentence and the final draft of the sentence one more time. You may experience two differences while reading the final draft of the sentence compared to the first: 1. Less stress and 2. More confidence that you understood what the writer wanted you to know. When you report results you have the same power to reduce emotional and cognitive strain for your reader by following the three recommendations for editing. These recommendations can be used when reporting any results, including the mean, median, and mode. When reporting results, be kind and helpful to your readers through editing.

Reading Review 3.3

Simplify each sentence using the three recommendations for editing.

1. The median GPA for a sample of students can be summarized as 2.80.
2. The GPA that was the most common in a data set was 3.00.
3. When an average GPA was computed by summing the sample GPAs and dividing it by the sample size, the average was 2.67.

3.6: Summary

After this chapter you should be able to do the following:

1. Find the mode of a dataset
2. Find the median of a dataset
3. Calculate the mean of a dataset
4. Calculate the deviation of individual scores from the mean
5. Use the sum of deviations to demonstrate the balance point of the mean
6. Compare the mode, median, and mean
7. Implement the three recommended considerations for editing results sentences

Structured Summary for Measures of Central Tendency

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes.

Summarize what each symbol stands for.

$n =$

$x =$

$\sum x =$

$\overline{x} =$

$\mu =$

$\tilde{x}$ or $\text{Mdn} =$

$Mo =$

Fill-in the appropriate information for each statistic below:

1. Mode
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What are the steps and/or formula and its steps for this statistic?
   d. How is this statistic reported when using APA format?
2. Median
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What are the steps and/or formula and its steps for this statistic?
   d. How is this statistic reported when using APA format?
3. Mean
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What are the steps and/or formula and its steps for this statistic?
   d. How is this statistic reported when using APA format?

3.7: Chapter References

American Psychological Association. (2022). *APA Style numbers and statistics guide*. https://apastyle.apa.org/ instructional-aids/numbers-statistics-guide.pdf

CHAPTER OVERVIEW

4: Measures of Variability

4.1: Variability

Statisticians use *measures of central tendency* and *measures of variability* to summarize variables in datasets. These two categories of measures focus on summarizing different aspects of the data in ways that can complement each other. The measures of central tendency summarize data by indicating where a set of scores tended to fall (see Chapter 3 for a full review of these measures). The mean is the most commonly used measure of central tendency for summarizing data from quantitative data measured using interval or ratio scales. However, measures of central tendency, such as the mean, do not provide information about how similar scores were to each other nor how similar scores were to the mean. Thus, they are more useful in meeting the goal of simplicity rather than specificity when summarizing data. Therefore, a separate but related set of techniques known as measures of variability are often used to bring some specificity into a summary.

Statisticians focus on measuring and analyzing data from variables, things which have the defining feature of variability. However, the measures of central tendency are limited by their inability to provide information about this important feature. Instead, measures of central tendency such as the mean are interpreted as describing what was generally true of the cases without being able to indicate the extent to which this tendency was *not* true. You may have noticed this limitation yourself when hearing descriptive summaries. For example, perhaps you have heard a professor say “The mean grade on the exam was 80 points.” You and a classmate might have looked at your own scores and found that neither of you had 80 points. Perhaps one of you had a score higher than 80 and the other had a score lower than 80 and you thought something along the lines of “Well, not everyone got an 80” or “The mean doesn’t represent my score.” This is the mean doing its job; it provides a simple summary of something that tended to be true but was not always true or exactly true for each case. If the mean was true of all cases, then the thing measured would be a constant and not a variable. Thus, it is best to think of any summary from a measure of central tendency as what was generally or approximately true but not what was exactly or always true.

The *measures of variability* bring specificity in by summarizing the dispersion of data. Dispersion focuses on the extent to which a group of scores tended to be similar (less disperse) or different (more disperse) from each other. Thus, **variability** in statistics refers to the dispersion or differences among scores or qualitative responses in a data set. Measures of variability are particularly useful when dealing with quantitative data. Quantitative data are often expected to follow patterns. Though there are options for summarizing variability for qualitative variables (such as the index of qualitative variation; IQV), the focus of this chapter will be four commonly used measures of variability for quantitative variables: exclusive range, inclusive range, standard deviation, and variance.

4.2: Range

We can summarize quantitative data using one of several versions of a measure of variability called the **range**. The way the range is computed depends on the nature of the variable and the intended use of the range.

The Exclusive Range

When the quantitative distance between the midpoints of the highest and lowest scores is desired, the exclusive range is used. This is a common version of the range which is often the one which is presumed used when a range is reported. The **exclusive range** refers to the difference between the highest value and the lowest value in a set of scores measured on an interval or ratio scale. These can be referred to as the maximum (or "max") and minimum (or "min"), respectively.

The formula has two parts:

1. X_{max} which represents the highest value raw score for the variable and
2. X_{min} which represents the lowest value raw score for the variable.

There is only one step to the formula which is to subtract because the result of subtraction is a *difference*.

Look over the ages in Data Set 4.1 and try to deduce the exclusive range. Note that, unless otherwise specified, you should assume that all data sets were generated from samples rather than populations. For the age data in Data Set 4.1, the exclusive range is 33. This is found by subtracting 14 (the X_{min}) from 47 (the X_{max}).

Data Set 4.1

Age
47
46
42
39
36
34
33
33
32
29
29
29
28
27
25
23
20
19
19
18
16
14

Exclusive Range

Formula	Calculation
$\mathrm{ER}=X_{\max}-X_{\min}$	$ER = 47 - 14$ $ER = 33$

The Inclusive Range

The **inclusive rage** refers to the difference between the highest value and one less than the lowest value in a set of scores measured on a quantitative scale of measurement. This version is generally useful when working with discrete values rather than continuous and can be applied for many forms of ordinal data as well as with data measured on the ratio or interval scales. This is because the inclusive range emphasized how many numbers are in the range, *including* the lowest value observed among the raw scores. Compare this to the exclusive range which *excludes* the lowest raw score and everything below it by subtracting it. In contrast, the inclusive range is only trying to remove everything below the lowest raw score while retaining the lowest score. Therefore, inclusive range is always one unit higher than the exclusive range.

The inclusive range is essentially showing how many numbers exist in the range. It could be applied to the data for ages for which the inclusive range would be 34. However, it is rarely used for a variable like this. Instead, you may see it used to identify how many score categories were in the range observed. Consider a quiz with 10 questions, each worth 1 point if correct and 0 if incorrect. In this example, there are 11 possible total scores an individual can get on the quiz: 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 and 0. If someone in the sample gets all 10 items correct (earning then 10 points) and another person get 0 items correct (earning them 0 points), the inclusive range will be 11 to represent all 11 possible scores in the range of observations, including the lowest score (which in this example is a 0).

Being specific about the exact statistic or formula used can reduce misconceptions and reduce the potential for unintentional misinformation. When the inclusive range is used it can be particularly important to state its full name because it is rarely used and, thus, readers may assume the more commonly used exclusive range is being reported when a statistician simple refers to their result as "the range." Different statistics are more or less common in different fields and specialties. Therefore, when a vague name is provided, a reader may incorrectly assume the term refers to the one they are most familiar with because it most easily comes to mind. In fact, we will see many terms, concepts, and statistics (both descriptive and inferential) with similar names. Thus, providing the full names for things helps ensure that information is accurately conveyed by minimizing opportunities for assumptions and misunderstanding.

Inclusive Range

Formula	Calculation
$\mathrm{IR}=X_{\max}-(X_{\min}-1)$	$IR = 47 - (14 - 1)$ $IR = 47 - 13$ $IR = 34$

> **Note**
>
> Here $X_{\max}$ refers to the highest raw score in the data set and $X_{\min}$ refers to the lowest.

Alternatively, some prefer to use this version of the inclusive range formula:

$$\mathrm{IR}=X_{\max}-X_{\min}+1$$

This is just another way to write the same formula which was shown before and both versions will always yield the same result. Some may prefer this version (which finds the exclusive range and then adds the lowest value back in via the + 1) because the order of operations is simpler. These two are both acceptable ways to compute IR. These versions of the inclusive range is appropriate for use with variables that are not quantitative and discontinuous (discrete) in nature.

The Real Limit Range

Another option for computing the range focuses on the real score limits. This is similar to the inclusive range but is a conceptually better fit for variables that are quantitative and continuous in nature. **Real limits** refer to the boundaries used to differentiate scores which are important for variables that are continuous in nature. When something continuous is measured, the precision of the quantity yielded is limited by the precision of the measurement instrument and operationalization of the variable. What essentially must happen when something continuous is measured is that it is cut off or rounded to some level of specificity. When something is truly continuous, the midpoints between possible scores yielded by the measurement tool used are the real score limits. For example, IQ scores are presented in whole numbers (rounded to the ones place) but the theoretical nature of intellectual functioning, if quantifiable, is presumed to be continuous. Therefore, an IQ test can yield scores of 99, 100, or 101, for example, each of which is actually functioning as a little interval rather than a single point on the number line. Theoretical IQ scores of 99.5 to $100.4\overline{9}$ (or 100.50 when rounded to the hundredths place) would all be represented as 100. Thus, the real lower limit of 100 is 99.5 and the real upper limit of 100 is $100.4\overline{9}$.

Visualizing Real Score Limits

*The real score limit of each value extends to the midpoint between it and each adjacent number (i.e. the values which would round to each whole number). This represents the real score limit range that each whole number represents when used to represent a continuous variable.

Raw Score	97	98	99	100	101	102	103
Real Score Limit of the Raw Score	96.50-$97.4\overline{9}$	97.50-$98.4\overline{9}$	98.50-$99.4\overline{9}$	99.50-$100.4\overline{9}$	100.50-$101.4\overline{9}$	101.50-$102.4\overline{9}$	102.50-$103.4\overline{9}$

We will refer to this version as the Real Limit Range to distinguish it from the exclusive range, though this is not a name that is often used in practice. The **real score limit range** refers to the difference between the upper-score limit of the highest value and the lower-score limit of lowest value in a set of scores measured on an interval or ratio scale when the variable is continuous in nature. These can be referred to as the upper real limit (URL) and lower real limit (or LRL), respectively.

The formula has two parts:

1. maxURL which represents the upper real limit of the highest value raw score for the variable and
2. minLRL which represents the lower real score limit of the lowest value raw score for the variable.

There is only one step to the formula which is to subtract because the result of subtraction is a *difference*.

This version of the range is operationally similar to the exclusive range (because each identifies boundaries of data and subtracts lower from higher) but yields results which are generally the same as the inclusive range. Essentially, the real score limit range is to data for variables which are presumed or known to be continuous what the inclusive range is to data for variables which are presumed or known to be discrete (discontinuous).

Review the dataset for the variable age (Data Set 4.1). The inclusive range is found by subtracting the lowest number in the set (which is 14) from the highest number in the set (which is 47) and then adding 1. For our age data the inclusive range is 34. Now look at the computations for the real score limit range (RSLR) and notice that, though they are conceptualized and computed differently, the RSLR and IR yielded the same result for Data Set 4.1.

Real Score Limit Range

Formula	Calculation
$\mathrm{RSLR} = X_{\max URL} - (X_{\min LRL})$	$\mathrm{RSLR} = 47.50 - 13.50$ $\mathrm{RSLR} = 34.00$

Note

Because decimals were introduced in the steps, it can be appropriate to round and show the answer to the hundredths place.

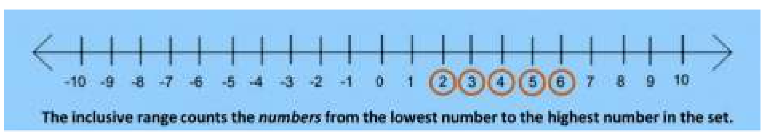
Comparing Ranges

Think about the difference between the inclusive range and exclusive range using the number line. The exclusive range refers to how many spaces we move on the number line to get from the lowest number in the dataset to the highest number. When we move on the number line, the exclusive range is counting how many spaces (also called "units") we have moved. The inclusive range, however, is asking how many numbers are included from the lowest number to the highest number. When we move on the number line, the inclusive range is counting each number we touch, rather than the spaces between them. Therefore, the inclusive range adds one so that it *includes* the lowest number and the highest number that exist in the dataset.

Let's take another example to help us understand the difference between exclusive and inclusive range. Suppose we gave a group of 10 students a 10-point quiz where the lowest score possible was 0 and the highest score possible was 10 (see the example data to the right). The scores in the dataset reflect the actual scores earned. The exclusive and inclusive ranges are each calculated using the scores earned, NOT the scores that were possible. Therefore, the range of possible scores, the exclusive range, and the inclusive range can all be different.

Quiz Scores	
5	6
3	6
4	2
2	5
3	6

The dataset includes scores of 2s, 3s, 4s, 5s, and 6s. Therefore, not all of the possible scores were earned. The exclusive range is calculated as: 6 – 2 = 4. This means that the highest score was four units from the lowest score. However, this does not reflect how many different scores were in the range because there were five different scores within the range (i.e. scores of 2, 3, 4, 5, and 6 are all in the inclusive range from 2 to 6). If we want to reflect how many numbers were in the range, we use the inclusive range. The inclusive range is calculated as: 6 – (2 –1) = 6 – 1 = 5. This means there were five numbers in the range, including the lowest and highest numbers in the range.



Reading Review 4.1

1. Which ranges can be used with quantitative data?
2. What distinguishes an exclusive range from an inclusive range?
3. How is the upper real score limit of a range defined?

4.3: Standard Deviation

Though the ranges help show how far scores could vary from each other, they do not indicate how far scores tended to be from a central score. This brings us to a measure of variability called the standard deviation. A **standard deviation** is a descriptive statistic that summarizes how far raw scores tended to fall from the mean in standard units. It can also be thought of as an estimate of the error that occurs when the mean is used to estimate the value of raw scores.

Interpreting the Standard Deviation

Standard deviations indicate how far scores tended to be from their sample mean. The smaller the standard deviation is, the closer the individual scores tended to be from the mean. The lowest a standard deviation can be is 0; when a standard deviation is 0 it means that scores did not deviate from the mean because they were all equal to the mean (and therefore are also all equal to each other). The larger the standard deviation, the farther the individual scores tended to be from the mean.

Standard deviation can also be interpreted as the expected error when using the mean to estimate a raw score from a sample. Means are used to summarize what tended to be true. Standard deviations summarize how far individual scores tended to be from the mean. Another way to say this is that the standard deviation estimates how wrong the mean is when used to represent the raw scores. Thus, this can be reworded to state that standard deviations summarize how far the mean is estimated or tended to be from raw scores. When the mean is more accurate in estimating the raw scores, the standard deviation will be lower. When the mean is perfect at estimating the raw scores because they are equal to the mean and each other, the standard deviation will be 0. When the mean is less accurate in estimating the raw scores, the standard deviation will be higher.

Notice that each of these two ways to interpret the standard deviation are simply different ways of saying the same thing: standard deviation estimates how similar or dissimilar scores tended to be from the mean and, thus, each other.

Calculating the Standard Deviation

There are two formulas for standard deviation: one for populations and one for samples. The formulas look complicated but actually use simple operations and can be quite easy to use once you have practiced them a few times. The core of each version of the formula is the same; standard deviation focuses on how far scores tend to deviate (i.e. be different) from the mean.

Decoding the Symbols

The symbol σ is used to refer to standard deviations of populations. The symbols *SD* or *s* are used to refer to the standard deviations of samples. x refers to an individual raw score. μ refers to the population mean, whereas $\bar{x}$refers to the sample mean. Remember, the mean is calculated the same way for populations and samples, even though the symbols are different. *N* refers to the size of a population, whereas *n* refers to the size of a sample.

Sum of Squares Within Formula

As is often the case in statistics, there are smaller formulas inside the standard deviation formulas. One which is subsumed into the standard deviation formula is known as the sum of squares within (often referred to simply as the "sum of squares" or SS). The **sum of squares within** is the sum of squared deviations from the mean within a sample. Its name is just a simplified version of its definition and operations. The sum of squares within can be represented using either the symbol *SS* or SS_w.

The *SS* formula and steps, in order, are summarized below.

Sum of Squares Within

Formula	Calculation Steps
$SS = \sum(x - \bar{x})^2$	1. Find the mean. 2. Subtract the mean from each raw score to find each deviation. 3. Square each deviation. 4. Sum the squared deviations.

The utility of this formula is not readily apparent on its own as it is rarely used on its own. Instead, it is best for summarizing some core steps that are built upon by other formulas such as the standard deviation and some inferential statistics such as ANOVA (which we will cover in Chapter 10). For now, we will focus on what the steps of the *SS* formula are doing.

The *SS* is a calculation of total deviations from the mean. However, the mean is a balance point of deviations. Therefore, if we were to just find each deviation and then sum, we would always get 0 for every data set. Therefore, a step is added where each deviation is squared to get rid of

all the negatives before summing. Thus, the *SS* reports the total squared deviations from the mean. In order to bring it from squared units back to standard units, we could add a step at the end where we square root the *SS* but you will notice this step is not included in the formula and you may wonder why. Recall that the *SS* is rarely used on its own and, instead, appears within other, larger formulas. Those formulas have more steps that will either include a step to square-root or some other procedures that require the sum of deviations be left in square units. Thus, the *SS* is left in squared units so that it is ready to go into other formulas that will deal with square rooting later on.

Population Standard Deviation Formula. You can see the population formula for standard deviation below. As we begin learning more complex formulas, it can be useful to notice their similarities and differences in structure relative to each other. There are a few core steps, components, and structural aspects of formulas that repeat in several different formulas in statistics. For example, you may notice a structural similarity between the standard deviation formula and the mean formula. Each includes a sum as a numerator and sample size as the denominator. The numerator of each formula sums whatever is the focus of the formula; thus, the numerator sums raw scores in the mean formula and sums deviations in the standard deviation formula. This is because the standard deviation uses the same logic as the mean to find how far scores tended to deviate from the mean. You can think of the standard deviation as the mean of deviations to help you remember what the formula is doing and understand why the formulas have similar structures. The standard deviation uses *SS* for its numerator because it allows the deviations to be summed without the negative deviations causing the sum to be 0. Thus, the first four steps of the standard deviation formula are the same as the steps for calculating *SS*.

The *SS* formula and steps, in order, are summarized below.

Standard Deviation for a Population

Formula	Calculation Steps
$\sigma=\sqrt{\frac{\sum(\mathrm{x}-\mu)^2}{N}}$	1. Find the mean. 2. Subtract the mean from each raw score to find each deviation. 3. Square each deviation. 4. Sum the squared deviations to find *SS*. 5. Divide *SS* by N. 6. Square root the result of step 5.

Notice that the last step is square rooting. This brings the calculation back into standard units which are easier to interpret than squared units and which are typically reported.

Sample Standard Deviation Formula

You can see the sample formula below. This formula is very similar to the population formula with two exceptions: 1. It uses sample symbols such as $\bar{x}$ in place of μ and 2. It uses an adjusted sample size in its denominator. This adjustment to the sample size is necessary because data for samples are incomplete and, thus, have more risk of error than data from populations. The standard deviation is an estimate of error so it is presumed to be higher when using samples. Decreasing the denominator, causes the quotient (i.e. the result from division) to increase. In keeping, the adjusted sample size is used to inflate the estimate of error when data are from samples rather than populations. Here are the steps of the sample standard deviation formula, in order:

Standard Deviation for a Sample

Formula	Calculation Steps
$s=\sqrt{\frac{\sum(\mathrm{x}-\bar{\mathrm{x}})^2}{n-1}}$	1. Find the mean. 2. Subtract the mean from each raw score to find each deviation. 3. Square each deviation. 4. Sum the squared deviations to find *SS*. 5. Find the adjusted sample size by subtracting 1 from n. 6. Divide *SS* by the adjusted sample size. 7. Square root the result of step 6.

Notice that the last step is square rooting to bring results back to standard units, just as it was for the population version of the formula.

> **Connecting Concepts**
>
> The formula used to find *s* includes a few parts. One important part of the formula looks like this "$\sum(x-\bar{\mathrm{x}})^2$" and is referred to as the *sum of squares* or with the symbol *SS*. Sum of squares is actually short for "the sum of squared deviations from the mean." Therefore, the s formula can also be written as:
>
> $$s=\sqrt{\frac{SS}{n-1}}$$

Sample Standard Deviation Formula Walkthrough

Let's work through an example of each step of the sample standard deviation formula using Data Set 4.1. One way to help organize the calculations is to use a table rather than writing them into a formula to make it easier to find and double-check each piece. Therefore, a table format will be used here to show the steps. Take a look at Table 1. The age data from Data Set 4.1 appear in the first column of Table 1 under the heading "Raw Score." Now that we have the data ready, we can start working through the steps of the formula.

1. **Find the Mean**.

We found that the mean was 29.00 for these data using the mean formula in Chapter 3 (see Chapter 3 to review the steps to calculate the mean, if needed).

2. **Find Each Deviation**.

The second step is to find the deviation for each raw score. This means that we must subtract the mean from each raw score. To make it easier to see and to organize this step, a second column appears in Table 1 which shows the sample mean to the right of each raw score. A third column appears to the right of that titled *Deviation*; this column shows what you would get when you subtracted the mean from each raw score. Notice that there are several positive deviations and several negative deviations and that, if you sum the deviations, the result will be 0. This will always be the case because the mean balances deviations and, thus, the sum of deviations from the mean must always be 0.

3. **Square Each Deviation**.

The third step is to square each deviation. This gets rid of the negatives seen in some of the deviations. The fourth column of Table 1 shows the squared deviation for each raw score.

4. **Sum the Squared Deviations**.

The fourth step is to add up, or sum, the squared deviations. This gives the Sum of Squares (or *SS*). You can see the sum of squared deviations is 1,850 in the bottom right corner of Table 1.

5. **Divide by the Adjusted Sample Size**.

The fifth step is to divide the sum of squares (*SS*) by the adjusted sample size. In step 4, we found that *SS* = 1,850. The adjusted sample size is calculated as *n* – 1. The sample size (n) for Data Set 4.1 is 22 making the adjusted sample size for this data set 21. To complete this step, we divide 1,850 by 21 as shown in the bottom of Table 1. The result, which appears under the square root sign in the formula, is 88.0952381…(remember that the ellipsis indicates that the number continues but is being abbreviated).

6. **Square Root**.

The sixth step is to square root to bring the value back into standard units. This step is necessary because everything was rounded in step 3 and we are now ready to put things back into standard units. When we square root 88.0952381… we get 9.3859063…, or 9.39 when rounded to the hundredths place. You can see this final step and the rounded answer in the bottom of Table 1.

Interpreting Standard Deviation

The standard deviation provides a summary of how different the raw scores tended to be from the mean. You can think of this as indicating how accurate the mean tends to be in describing each individual in the sample (or population). Remember, a sample mean provides a summary of the whole sample. Thus, the standard deviation is also a summary of the whole sample in relation to its mean. The standard deviation can be described as the average error we would have if we assumed each member of the sample had a score for the variable that was the same as the sample mean. Of course, the raw scores are often similar, but not identical, to the mean.

Remember, the smaller the standard deviation, the more similar sample raw scores tended to be to the mean. If the standard deviation was 0, it would indicate that there was no deviation. That is the same as saying every raw score was the exact same value as the mean and, thus, that the data were constant rather than varied. The larger the standard deviation, the less similar the sample raw scores tended to be to the mean. There are a few other ways we could convey this same concept. We could say that when standard deviations are smaller, there is less dispersion and that when they are larger, there is more dispersion. Another way we could interpret this is to say that the mean has less error in estimating raw scores when the standard deviation is small compared to when the standard deviation is large. As should be clear here, the standard deviation is a descriptive statistic that builds from and accompanies the mean in describing data.

Table 1 Standard Deviation Calculations for Data Set 4.1 (n = 22)

Raw Score	Mean	Deviation	Squared Deviation
47	29	18	324
46	29	17	289
42	29	13	169
39	29	10	100
36	29	7	49
34	29	5	25
33	29	4	16
33	29	4	16
32	29	3	9
29	29	0	0
29	29	0	0
29	29	0	0
28	29	-1	1
27	29	-2	4
25	29	-4	16
23	29	-6	36
20	29	-9	81
19	29	-10	100
19	29	-10	100
18	29	-11	121
16	29	-13	169
14	29	-15	225

Table 1 Standard Deviation Calculations for Data Set 4.1 (n = 22)

Sum of Deviations = 0	Sum of Squared Deviations (SS) = 1,850
$s=\sqrt{\frac{1,850}{22-1}}=\sqrt{\frac{1,850}{21}}=\sqrt{88.0952381\ldots}=9.3859063\ldots\approx 9.39$	

The Mean and Standard Deviation Go Together

In statistics and research, we typically report the mean and the standard deviation together because they complement each other in summarizing data. This is because they tell us a more complete story together than either does on its own. The mean tells where the group of scores tended to be. The standard deviation tells us how far raw scores tended to be from the mean. When the mean and standard deviation appear together, they provide summaries of data that aid in making useful comparisons in a way that neither can do alone. For example, you can have the same mean in two samples but different standard deviations. The sample with the smaller standard deviation tended to have raw scores that were closer to the mean than the sample with the larger standard deviation.

Let's review this concept with some examples. Assume five groups of individuals (i.e. five samples) reported how many hours they spent working per week. Assume also that the sample size of each of the five groups is the same. We can look at the means and standard deviations (*SD*s) for each group instead of all of the raw scores. This makes it easier to compare the groups because the mean and *SD* summarize each group in a comparable way.

Group 1: $M = 40.00$, $SD = 0.00$

Group 2: $M = 40.00$, $SD = 5.00$

Group 3: $M = 40.00$, $SD = 20.00$

Group 4: $M = 20.00$, $SD = 1.00$

Group 5: $M = 20.00$, $SD = 5.00$

Groups 1, 2, and 3 have the same mean but different *SD*s. The *SD* for Group 1 indicates that everyone worked for 40 hours because there is no deviation (i.e. no differences from the mean). The *SD* for Group 2 indicates that folks worked about 40 hours but that individuals varied. Some worked more, others less. The same is true of Group 3, however, the *SD* for group 3 is very large. This means that scores in Group 3 varied much more greatly around the mean than scores in Group 2. Compare the means and *SD*s in groups 4 and 5. You should notice that the raw scores in Group 4, on average, were mores similar to the mean than the raw scores in Group 5 were, on average.

Remember, variables are things that vary. The *SD* is a way to summarize how greatly raw scores of a variable tended to vary from their group's mean. The larger the *SD*, the greater the dispersion. The smaller the *SD*, the lesser the dispersion.

4.4: Summarizing Results in Sentences

If the *SD* for the age variable in Data Set 4.1 was 9.39 when rounded to the hundredths place. Let's now focus on how to report results like these in American Psychological Association (APA) format. When summarizing Data Set 4.1, one could write:

> Data were collected from 22 people. They ranged in age from 14 to 47 (*M* = 29.00, *SD* = 9.39).

The summary sentences use a format referred to as APA. In APA format, symbols can be written in parentheses to provide important results in a concise way. When a mean is reported in parenthesis, the symbol *M* is used. When a standard deviation is reported in parenthesis, the symbol *SD* is typically used. The summary sentences clearly and succinctly give a lot of information about the data to the audience. The audience knows the sample size. The audience knows that the lowest age was 14 and the highest age was 47. The audience also knows that the mean score was 29.00 but that individual ages varied around the mean quite a bit. By always presenting the mean and standard deviation together we get a nice snapshot. Image you were reading a paper and it contained this sentence:

> Data were collected from 50 people. The mean age in the sample was 15.00 (*SD* = 0.00).

What do we know about this dataset? Well, the standard deviation reflects how far people tend to fall from the mean and it is reported as 0.00. Thus, everyone was 15 years of age! Remember, when a standard deviation is 0.00 it means that everyone had the exact same score and there is no variation in the data.

Examples of APA Summaries for Data Set 4.1

There are several ways that data for a variable can be summarized using APA format. Here are a few examples of different ways to summarize the same data:

Option 1: Data were collected from 22 people. They ranged in age from 14 to 47 (*M* = 29.00, *SD* = 9.39).

Option 2: Data were collected about age (n = 22). The exclusive range for age was 33.00 (*M* = 29.00, *SD* = 9.39).

Option 3: Age varied (*M* = 29.00, *SD* = 9.39) in a sample of 22 persons.

Often sample size is already reported elsewhere in a report, in which case a simpler sentence focused on the mean and standard deviation can be used like so:

Option 4: The mean age was 29.00 (*SD* = 9.39).

Option 5: The mean age was 29.00 with a standard deviation of 9.39.

The symbols M and SD are used to represent the mean and standard deviation, respectively, when these are presented in parenthesis as part of a sentence. Sample size is represented with the symbol n.

Reading Review 4.2

1. What symbols should be used for the mean and standard deviation when they are reported within parentheses in a sentence?
2. What does the standard deviation add to the summary which is why it is usually reported with the mean?
3. Do the symbols for mean and standard deviation always need to be used when summarizing each in sentences?

4.5: Variance

Another important measure of variability is the variance. **Variance** is a descriptive statistic that summarizes how far raw scores tended to fall from the mean in squared units. It can also be thought of as the squared estimate of the error that occurs when the mean is used to estimate the value of a raw score. Thus, the variance is just the standard deviation left in squared units.

However, we generally best understand and, thus, report results in standard units. For this reason, variances are rarely reported as final results. Instead, variances are often used as important parts of other formulas (such as in an independent samples *t*-test which we will cover in Chapter 8).

Decoding the Symbols

The symbols within the variance formulas are the same as those within the respective standard deviation formulas: x refers to an individual raw score, μ refers to the population mean, $\bar{x}$ refers to the sample mean, N refers to the population size, and n refers to the sample size. The variance is just the standard deviation left in squared units. For the sake of simplicity and to keep this clear, the symbols for variance are just the standard deviation symbols adjusted to reflect this. Specifically, σ^2 is used to refer to variances of populations and SD^2 or s^2 are used to refer to the variances of samples.

Calculating the Variance

Just as was true for standard deviation, there are two formulas for variance: one for populations and one for samples. The formulas use the same symbols and many of the same steps as the standard deviation.

Population Variance

Here is the formula and required steps for finding the variance of a population:

Variance for a Population

Formula	Calculation Steps
$\sigma^2=\frac{\sum(x-\mu)^2}{N}$	1. Find the mean. 2. Subtract the mean from each raw score to find each deviation. 3. Square each deviation. 4. Sum the squared deviations to find SS. 5. Divide SS by N.

Notice that these are the same as the first five steps to the population standard deviation formula. This leaves the result for variance in squared units, as is indicated by the symbol for variance which is σ^2.

Sample Variance Formula

You can see the sample variance formula below. This formula is very similar to the population formula with two exceptions:

1. It uses sample symbols such as $\bar{x}$ in place of μ and
2. It uses an adjusted sample size in its denominator.

These are the same exceptions we saw when we moved from the formula for population standard deviations to the formula for sample standard deviations and they serve the same purpose here.

Variance for a Sample

Formula	Calculation Steps
$s^2=\frac{\sum(\mathrm{x}-\bar{\mathrm{x}})^2}{n-1}$	1. Find the mean. 2. Subtract the mean from each raw score to find each deviation. 3. Square each deviation. 4. Sum the squared deviations to find *SS*. 5. Find the adjusted sample size by subtracting 1 from n. 6. Divide *SS* by the adjusted sample size.

Note that because the variance is calculated using the first six steps to finding a standard deviation, we have already calculated the variance for Data Set 4.1 when we found its standard deviation. Take a look at Table 2 which shows calculations for variation. Everything is the same as in Table 1 which showed the calculations for standard deviation for Data Set 4.1 until the last step. Specifically, instead of using the symbol s for standard deviation the symbol s^2 for variance appears and the square root sign is missing. Therefore, the result, when rounded to the hundredths place, is simply 88.10. This is the same value we had under the square root sign in the penultimate step for calculating standard deviation in Table 4.1. By comparing the two tables we can see how the variance was found in Table 4.1 in the process of ultimately calculating the standard deviation.

Table 2 Standard Deviation Calculations for Data Set 4.1 (n = 22)

Raw Score	Mean	Deviation	Squared Deviation
36	29	7	49
34	29	5	25
33	29	4	16
33	29	4	16
32	29	3	9
29	29	0	0
29	29	0	0
29	29	0	0
28	29	-1	1
27	29	-2	4
25	29	-4	16
23	29	-6	36
20	29	-9	81
19	29	-10	100
19	29	-10	100
18	29	-11	121
16	29	-13	169
14	29	-15	225
		Sum of Squared	
		Sum of Deviations = 0	Deviations (SS) = 1,850

$$s^2 = \frac{1,850}{22-1} = \frac{1,850}{21} = 88.0952381\ldots \approx 88.10$$

4.6: Choosing a Measure of Variability

The ranges, standard deviation, and variance are all options for summarizing variability of quantitative variables. Each differs in its utility and, thus, when it is most likely to be used.

Utility of Ranges

The ranges are the most basic and easiest forms of variability to calculate. They are best used when the aspect of variability of interest is simply how wide the distance between observed scores was. Typically, this is used when the mean is not of interest but may be used when a mode or median is of interest as a way to summarize central tendency. However, ranges can also be reported on their own without a measure of central tendency.

Utility of Standard Deviation

The standard deviation is the most complex measure of variability to compute of the options reviewed here. However, it is often used and is the measure of variability of choice to report when a mean is also being considered and/or reported. It adds important information about how close or far raw scores tended to fall from the mean. Where the mean offers simplicity in summarizing a score, the standard deviation adds specificity to the summary. Therefore, the standard deviation and mean are often reported together to provide a useful and balanced summary of a quantitative variable.

Utility of Variance

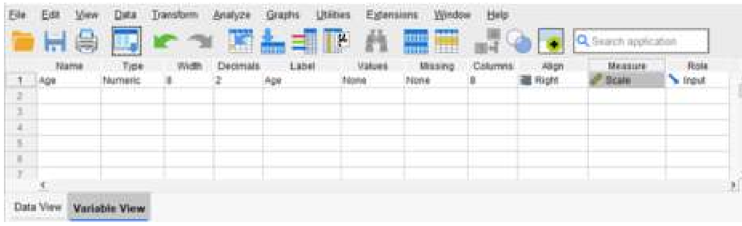
The variance is rarely reported because it is in squared units. Therefore, the standard deviation is often reported alongside the mean but the variance is not. This leaves us with the question of when variance would be used and why it matters. The main way we will see variance used is within other formulas when the concept of standard deviation is being used but the formula has other work to do before square rooting. Thus, the variance is used inside of other formulas which consider multiple things together which will be square rooted later. Basically, when that happens, we are saying we want to use the concepts of the standard deviation but are not ready to square root just yet. We will see this use of variance when we learn about the independent samples *t*-test in Chapter 8.

Reading Review 4.3

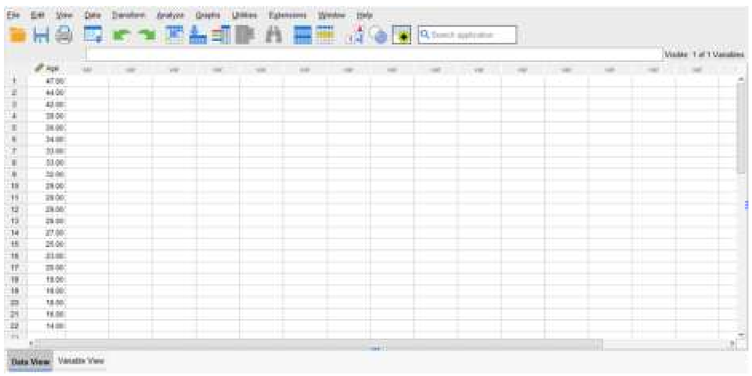
1. What distinguishes a variance from a standard deviation?
2. If a variance is known, what can be done to it to find its corresponding standard deviation?
3. What is the utility of the variance?

4.7: Using SPSS

SPSS software can be used to find the standard deviation and variance of a variable. Before analyzing data, they must be properly entered and defined in an SPSS spreadsheet. To review the process for entering data, see Chapter 2. If data are entered for Data Set 4.1, the variable view tab should look like this:



If data are entered for Data Set 4.1, the data view tab should look like this:

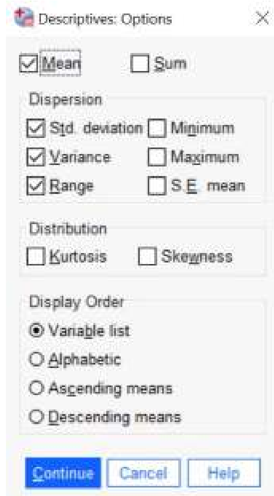


Once data are properly entered, the steps below can be followed.

Analyzing Data

The steps to generating a standard deviation in SPSS are:

1. Click Analyze › Descriptive Statistics › Descriptives.
2. Drag the name of the variable you want to analyze from the variable list on the left into the Variable(s) text box on the right of the command window. You can also do this by clicking on the variable name to highlight it and the clicking the arrow to move the variable from the left into the Variable text box on the right.
3. Click the Options button on the upper right side of the command window.
    a. Some items will be marked by default such as the mean. Click on the boxes to select any measures of variability desired such as the range, standard deviation, and variance. You can also select boxes to have SPSS specify what the highest and lowest scores were for the variable, if desired. The image to the right is what the Options window would look like if requesting the mean, range, standard deviation, and variance. Click Continue to close the options window and return to the main command window.
4. Click OK.



5. A table summarizing the results will appear in the SPSS output viewer. Here is an example of what that would look like for Data Set 4.1 for the items selected as shown in step 3 above:

Descriptive Statistics

	N	Range	Mean	Std. Deviation	Variance
Age	22	33.00000	29.00000	9.38591	88.09524
Valid N (listwise)	22				

Compare the results in the SPSS output table to the ones computed by hand in this chapter. The sample size is 22, which matches Data Set 4.1. Only one range is reported in SPSS which is the exclusive range, however, it is important to note that SPSS does not specify and simply refers to this as "Range." The range as reported in the SPSS output table matches the exclusive range computed by hand in this chapter as 33.00. SPSS also reports the mean as 29.00, the standard deviation as 9.39, and the variance as 88.10 when rounded to the hundredths place. These are consistent with the values computed by hand in this chapter. By comparing our hand-calculated results to those provided by SPSS we can get assurance that our computations are correct.

Structured Summary for Measures of Variability

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes.

Summarize what each symbol stands for and the types of data for which each is most appropriate.

IR =

ER =

$RSLR$ =

n =

N =

x =

$\bar{x}$=

μ =

SS =

s =

σ =

s^2 =

$\sigma^2 =$

Fill-in the appropriate information for each statistic below:

1. Exclusive Range
    a. For which kinds of data can/should this be used?
    b. What is the focus of this statistic?
    c. What are the steps and/or formula and its steps for this statistic?
2. Inclusive Range
    a. For which kinds of data can/should this be used?
    b. What is the focus of this statistic?
    c. What are the steps and/or formula and its steps for this statistic?
3. Standard Deviation
    a. For which kinds of data can/should this be used?
    b. What is the focus of this statistic?
    c. What are the steps and/or formula and its steps for this statistic?
    d. How is this statistic reported when using APA format?
4. Exclusive Range
    a. For which kinds of data can/should this be used?
    b. What is the focus of this statistic?
    c. What are the steps and/or formula and its steps for this statistic?

CHAPTER OVERVIEW

5: The Normal Curve, z-Scores, and Probability

5.1: The Normal Curve, z-Scores, and Probability

As we have reviewed in the prior two chapters, the mean and standard deviation are, together, often the foundation used to summarize and describe a quantitative variable. In this unit, we will build on our knowledge of the mean and standard deviation by exploring how they relate to a graph known as the normal distribution curve. We will also learn how z-scores are computed using means and standard deviations and how they describe the positions and probabilities of raw scores on the normal distribution curve.

Descriptive statistics such as the mode, median, mean, and standard deviation are associated with probabilities. This is because these four descriptive measures exist in meaningful places on something known as the normal distribution curve (see Figure 1). The normal distribution curve is also sometimes referred to as the normal curve for short, as the bell curve due to its bell-like shape, or as the Gaussian distribution named for a scientist and mathematician who wrote using the concepts of the distribution to describe patterns of measurement error. We can use our knowledge about these descriptive statistics and the curve to better understand our sample data and to begin to make inferences about the populations they represent.

5.2: The Normal Distribution Curve

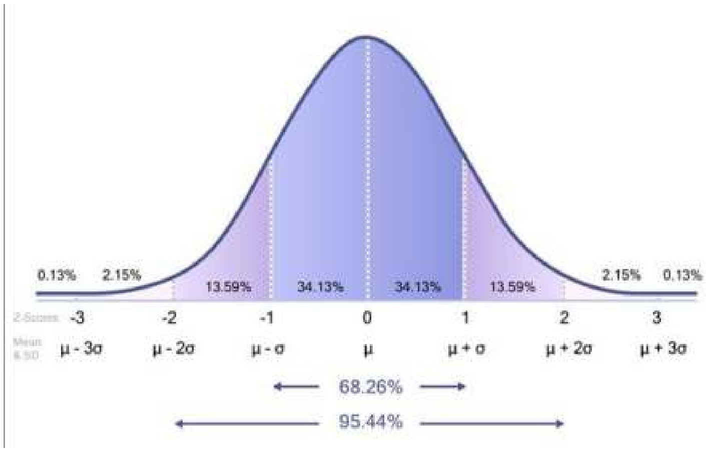


Figure 1.

You might have heard someone refer to data as “normal” before and assumed it meant nothing was unusual or problematic about the data. This is partially true but we must distinguish between what the word “normal” means in general compared to in the field of statistics, specifically. In general, when people say normal, they often mean some version of “This is as expected” or sometimes they mean “This isn’t weird.” The latter can have a negative tone when someone is using the word normal to mean good and abnormal to mean bad. However, for a statistician, normal and abnormal are not synonymous with good or bad. Instead, in statistics, saying data are normal means that the data follow a specific, expected pattern that is visually represented by the normal distribution.

The normal distribution is a probability graph which is commonly referred to in statistics. A probability graph is one which is used to represent how common (likely) or rare (unlikely) various scores are. You will notice that the word “scores” was used; this is because the normal distribution is used to represent the probabilities of various scores or score intervals for quantitative variables rather than qualitative variables. This graph is created univariately meaning only one variable is expected to be represented with the graph at a time, though multiple normal curves can be shown together to allow for comparisons of different variables or groups (which we will see in our later chapters about independent samples *t*-tests and ANOVA).

Key Features of the Normal Curve

There are key features of the normal distribution that make it easy to visually distinguish from other graphs: the peak, asymptotic nature, and symmetry of the graph. The normal curve is essentially a frequency polygon which is tallest (peaks) at the center and gets progressively shorter as you move further into the tails. The word “tails” refers to the outer portions of the graph where the curve takes on a noticeably flatter (more horizontal) appearance or slope. The height of the curve corresponds to the y-axis and represents the frequency with which scores occurred. This means that scores at the middle of the distribution (represented as the apex of the curve) are those that are most frequently occurring. This also means that scores become less common as we move farther from the middle in either direction. Another way to say this is that scores are more common in the tall parts of the graph and less common in the shorter (tail) parts of the graph. When data are perfectly normally distributed, the mode, median, and mean will be the same number and will be on the x-axis directly under the apex or peak of the curve. The normal distribution is also symmetrical. This means that the left side of the graph is a mirror image of the right side of the graph.

The Structure of the Normal Curve

The Axes. The x-axis of the normal curve is used to show quantitative raw scores going from lower scores on the left to higher scores on the right. Because the graph is asymptotic, it theoretically extends to infinity in both directions, however, it would be impossible to actually draw the graph out to infinity. Therefore, the graph usually depicted extending only far enough to show that the tails are getting quite close to the x-axis. Essentially, then, the section of x-axis that is usually shown represents the most relevant segment of a number line which extends from -3 to 3. The y-axis is used to represent the frequencies with which each score occurred with the theoretically lowest possible y-value being 0 to represent the absence of a particular score. These are the same things which are represented on the x-axes and y-axes of histograms and frequency polygons (which were covered in Chapter 2). In fact, you might notice that the normal curve is actually just a very smooth frequency polygon!

The Position of the Mean and Standard Deviation

The mean and standard deviation are used to define and differentiate areas of the normal curve. The mean falls on the x-axis directly under the center, or peak, of the curve, causing half of the graph to be to the left of the mean and the other half to be to the right of the mean. This means that, proportionally, half of the raw scores are lower than the value of the mean (corresponding to the area to the left of the curve's center) and half of the raw scores are greater than the value of the mean (corresponding to the area to the right of the curve's center).

The standard deviation (*SD*) is then used to distinguish various locations to the left or right of the center of the normal curve. Recall that the standard deviation refers to how far individual raw scores within a sample tended to fall from their sample mean (see Chapter 4 for review). Thus, the standard deviation is a way of describing how much individual scores for a variable

tended to differ from the mean. This is the same way it is used in the normal curve. The mean is the center and standard deviations are subtracted from the mean as we move to the left (meaning we are moving to scores that deviate increasingly more by being lesser than the mean) and added to the mean as we move to the right (meaning we are moving to scores that deviate increasing more by being greater than the mean).

Seven specific locations are generally marked along the x-axis using the mean and *SD*. These locations, from left to right, represent scores that are 3 *SD*s below the mean, 2 *SD*s below the mean, 1 *SD* below the mean, at the mean (i.e. 0 *SD*s away from the mean), 1 *SD* above the mean, 2 *SD*s above the mean, and 3 *SD*s above the mean. Scores which are more than 3 standard deviations from the mean are extremely rare in the normal curve and, thus, it is often sufficient to only show these 7 specific locations to accommodate the vast majority raw scores on the x-axis.

These seven locations offer convenient markers to help us see that scores at the mean are most common and that scores become less common the more standard deviations they are away from the mean. For example, when data are normally distributed, scores at the mean are the most frequently occurring. A raw score that is one standard deviation lower than the mean is less frequently occurring than a raw score which is at the mean; however, a raw score that is one standard deviation below the mean is just as frequently occurring as a raw score which is one standard deviation above the mean. This is because the decrease in height on each side of the graph is symmetrical to the other. Further, a raw score which is two standard deviations below the mean is more common than a raw score that is three standard deviations below the mean, is just as common as a raw score which is two standard deviations above the mean, and is less common that raw scores that are one standard deviation below the mean, those which are at the mean, and those which are one standard deviation above the mean.

Reading Review 5.1

Try to answer the following questions regarding the Normal Curve:

1. What are the three distinguishing features of the Normal Curve?
2. What is represented on the y-axis of the Normal Curve?
3. What is represented on the x-axis of the Normal Curve?
4. Where is the mean found in the Normal Curve?
5. What are the seven positions which are usually shown on the x-axis of the Normal Curve?

Using z-Scores for the x-Axis

You may have noticed that this section focused heavily on where the mean and standard deviations are in the normal curve and how to identify those locations. Because of this focus, statisticians often use something known as z-scores to represent the locations of the mean and the number of standard deviations various raw scores are from the mean.

Raw scores are often converted to z-scores to make their location on the x-axis and their corresponding probabilities easier to understand and compare. A z-score tells us how many standard deviations a raw score is from the sample mean. If a raw score is equal to the mean it does not deviate from the mean at all and, thus, its z-score would be equal to 0. This raw score would be at the center of the curve. If you look at Figure 2, you will see that there are two rows of labels under the x-axis: one refers to the location of the mean and raw scores and the other refers to their corresponding z-scores. Thus, z-scores align with, and indicate, how many standard deviations each raw score is from the mean. Notice how much easier it is to say "corresponds to the z-score of 1" compared to how cumbersome it is to say "corresponds to a raw score which is one standard deviation above the mean." These actually refer to the same locations under the curve yet using z-scores to identify locations is much simpler. For this, and other reasons, statisticians prefer to use z-scores to refer to locations under the curve.

Take a moment to compare the two rows of information under the x-axis to familiarize yourself with what z-scores indicate. When a score is equal to the mean, it is at the center and labelled as a z-score of 0. When a score is below the mean, it is to the left and gets a negative z-score. The negative sign indicates that the corresponding raw score for that location is lower in value than the mean. The value of the z-score indicates how many standard deviations the score is from the mean. Therefore, when z = -1 it refers to where a raw score which is one standard deviation less than the mean is located in the normal curve. When z = -2 it refers to where a raw score which is two standard deviations less than the mean is located in the normal curve. Consistent with this, when a score is above the mean, it is to the right and gets either a positive sign or no sign (because the absence of a sign also indicates that a value is positive in mathematics). Therefore, when z = 1 it refers to where a raw score which is one standard deviation greater than the mean is located in the normal curve. When z = 2 it refers to where a raw score which is two standard deviations greater than the mean is located in the normal curve, and so on.

Remember that the normal curve is just a polygon for a population histogram and, thus, the height of the line reflects frequencies of occurrences at points along the x-axis. In the normal distribution, scores nearest the mean are more common and occurrences are expected to get less and less common the further we move away from the mean along the x-axis. This corresponds to the probability that each person who is randomly selected will have a score at various places along the x-axis.

Probability and the Normal Curve

The area under the curve at any given place or section is equal to the probability that a score is in that area. Therefore, the likelihoods that scores exist is computed using their corresponding areas under the curve. Some of these probabilities are easy to see while others are a bit more challenging. For example, we know that the mean, median, and mode are all at the center which divides the graph symmetrically in half. Therefore, we know that half of the raw scores are expected to be to the left of mean and the other half are expected to be to the right of the mean. Because the tails of the graph are asymptotic, meaning they get progressively closer to the x-axis without ever touching or crossing it, we know that 100% of scores are expected to be somewhere under the curve. Think about it this way: If the tails are asymptotes that, theoretically, can continue out to negative and positive infinity, every value of the number line is being represented on the x-axis. Therefore, every possible value of X (and every possible case when measuring X) is somewhere in the curve. This would mean that 100% of all values of X are somewhere in the curve. However, finding the probabilities of other segments of the curve requires us to consider the sloping nature of the curve, the role of the standard deviation, and/or the corresponding z scores.

The standard deviation is an important part of the normal distribution. When data are normally distributed, 34.13% of cases are expected to fall within one standard deviation below the mean to the mean. When data are normally distributed, 34.13% of cases are expected to fall within one standard deviation above the mean to the mean. This means that 68.26% of the population is expected to be within one standard deviation of the mean (half of whom are expected to be lower than the mean an half of whom are expected to be higher than the mean; see Figure 2).

What happens to the scores that are at the mean?

These are theoretically divided in half such that half of the scores at the mean are treated as being very slightly below the mean and half are treated as being very slightly above the mean when estimating the proportions of scores to the left and right of the mean line.

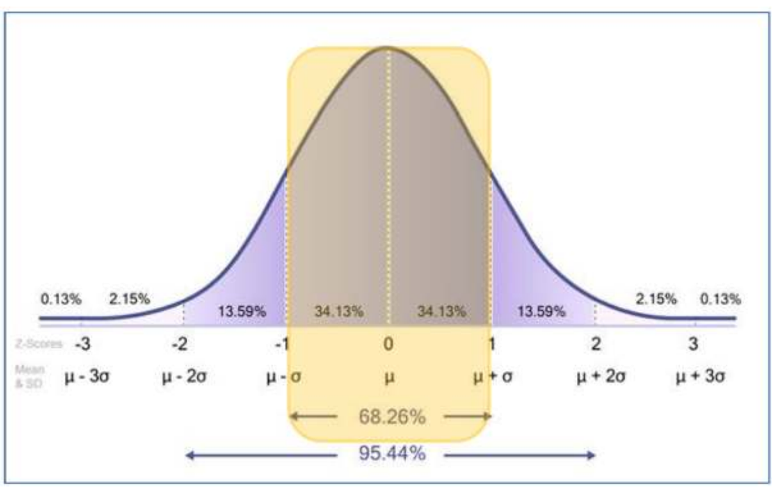


Figure 2.

You can see these two portions of the curve highlighted by the yellow box in Figure 2. If you were to add those two areas within one standard deviation below the mean and one standard deviation above the mean you would get the total of 68.26% that you see reflected on the graph. Note that in the graph, the percentages are rounded to two decimal places.

Reading Review 5.2

Try to answer the following questions using the graph:

1. What proportion of cases are expected to fall below the mean?
2. What proportion of cases are expected to fall above the mean?
3. What proportion of cases are expected to fall between one and two standard deviations below the mean?
4. What proportion of cases are expected to fall between one and two standard deviations above the mean?

5.3: z-Scores

A z-score is used to summarize how far an individual case is from the mean in standard deviations. The z-score is often also referred to as the standard score. Each z-score indicates two important things about the location of a raw score: its direction and distance from the mean.

Direction

The sign of a z-score indicates its direction. When a raw score does not deviate from the mean, it is equal to the mean. However, raw scores can also deviate by being below the mean or above the mean. When z = 0, it means the raw score did not deviate from the mean and is, thus, equal to the mean. Keep in mind that 0 is not considered a signed number because it represents the dividing point between signed numbers (which are either positive or negative). When a z-score is positive, it means that the raw score was greater than the mean. When a z score is negative, it means a raw score was lesser than the mean. Therefore, by looking at the sign of any z-score, you can instantly deduce whether the raw score was greater, less than, or equal to the mean.

Distance

The size of the z-score indicates distance from the mean. Because z-scores are computed to show how many standard deviations a raw score is from a mean, the value of the z-score can be used to quickly deduce how far a raw score is from a mean. The smaller the absolute value of a z-score, the closer a raw score is to the mean. Conversely, the larger the absolute value of a z score, the further a raw score is to the mean. When a raw score is equal to the mean, z = 0; zero represents the absence of distance from the mean.

z-Score Formula

Because z-scores are quite useful, raw scores are often converted to z-scores. This is done using a simple two step formula:

Z-Score Computations

Formula	Calculation Steps
$z=\dfrac{x-\mu}{\sigma}$	1. Subtract the mean from a given raw score to find the deviation. 2. Divide the deviation by the standard deviation.

Notice that the z-score formula uses population symbols. This is the standard way to write the formula because the normal curve is focused on broader truths that are presumed to apply to populations. However, sometimes the formula is written with sample symbols when working with sample data. The steps are the same regardless of whether the sample or population symbols are used. The z-score formula written with the sample symbols looks like this:

$$z=\frac{\mathrm{x}-\bar{x}}{\mathrm{s}}$$

The formula requires knowing the mean and standard deviation for a set of scores before plugging those in to convert any given raw score to its z-score. If the mean and standard deviation are already known, they can be plugged in with a given raw score and used to evaluate z. When we say we are “evaluating” with a formula, it means we are simply computing a result by following the steps of a formula or algebraic expression.

Finding z with a Given M and SD

Let’s try with an example. Suppose data were gathered on the amount of coffee beans customers purchased measured in ounces. The quantitative variable being measured is *weight*. The unit of measurement is *ounces*. The sample is *bags of coffee sold*. Weight is our X variable and each individual raw score is an example of X. Suppose that when the data were summarized, it was found that the mean was 16.00 ounces with a standard deviation of 4.00. These summaries are commonly provided in APA-formatted sentences like this:

The mean weight for the sample of coffee bags was 16.00 ounces (*SD* = 4.00)

Now that we are given the mean and standard deviation, we can convert any given raw score to its corresponding z-score. Suppose we want to know the z-score for the first bag of coffee sold and that this bag weighed 12.00 ounces. We could summarize this by saying X_1 = 12.00. You can read this as saying “The raw score of X for case 1 was 12.00.” Now we have all that we need to find the z-score for the raw score of 12.00. Let’s organize our information to get ready for using the formula:

$\bar{x}$= 16.00

s = 4.00

X_1 = 12.00

Now we can plug these into the z-score formula to find z. We will use the sample symbols to remind ourselves that we are estimating using sample information rather than known population values.

> Tip
>
> Always start by writing the formula down before plugging in any values. This helps you get familiar with the formula and its parts. This can help you to both recognize and memorize the formulas.
>
> $$z=\frac{\mathrm{x}-\bar{x}}{s}$$
> $$z=\frac{12.00-16.00}{4.00}$$
> $$z=\frac{-4.00}{4.00}$$
> $$z=-1.00$$

We can follow this same process to convert any raw score to a z-score. Here are three examples of different raw scores for weights of coffee bags being converted; this example is focused on a sample and, thus, sample symbols are used:

Example 1	Example 2	Example 3
Given: $\bar{x}$= 16.00 s = 4.00 X_1 = 21.00 Find z_1: $z=\frac{\mathrm{x}-\bar{x}}{s}$ $z=\frac{21.00-16.00}{4.00}$ $z=\frac{5.00}{4.00}$ $z=1.25$	Given: $\bar{x}$= 16.00 s = 4.00 X_2 = 8.80 Find z_2: $z=\frac{\mathrm{x}-\bar{x}}{s}$ $z=\frac{8.80-16.00}{4.00}$ $z=\frac{-7.20}{4.00}$ $z=-1.80$	Given: $\bar{x}$= 16.00 s = 4.00 X_3 = 16.40 Find z_3: $z=\frac{\mathrm{x}-\bar{x}}{s}$ $z=\frac{16.40-16.00}{4.00}$ $z=\frac{0.40}{4.00}$ $z=0.10$

Finding z Starting with Raw Data.

Sometimes raw data are provided which require the statistician to do a little more work before using the formula. When raw scores are provided, the statistician must first find the mean and standard deviation before plugging them in to solve for the z-score of a given raw score.

Let's take a look at Data Set 5.1; this data set is for the variable *volume of coffee consumed* measured in ounces for a sample of 22 customers. First, we can start by finding sample size, then the mean, and finally, the standard deviation. A summary of these calculations is shown at the bottom of Data Set 5.1. The standard deviation is shown rounded to the fourth decimal place (also known as rounded to the ten-thousandths place). For a detailed review of how to calculate the mean, see Chapter 3. For a detailed review of how to calculate the standard deviation, see Chapter 4.

Once the mean and standard deviation have been computed for a sample, the z-score formula can be used to calculate the z-score for any raw score. Let's find the z-score for the first person in the data set who had a raw score of 12. Thus, X_1 = 12.00.

Data Set 5.1

Volume of Coffee Consumed in Ounces	
12	8

Volume of Coffee Consumed in Ounces	
11	8
10	7
10	7
9	7
9	7
9	7
9	6
9	6
8	5
8	4
Descriptive Statistics n = 22 $\bar{x}$= 8.00 s ≈1.9024	

Calculations with Data Set 5.1

Given:

$\bar{x}$= 8.00

s = 1.9024

X_1 = 12.00

Find z_1:

$$z = \frac{\mathrm{x} - \bar{x}}{s}$$
$$z = \frac{12.00 - 8.00}{1.9024}$$
$$z = \frac{4.00}{1.9024}$$
$$z \approx 2.1026$$

The z-score for X_1 is 2.10 when rounded to the hundredths place. It can be summarized as: Z_1 = 2.10

Thus, the only difference in finding a z-score when given raw data is that you must first calculate the mean and standard deviation. After that, values can be plugged into the z-score formula to solve for z.

Converting from z-Scores to Raw Scores

Sometimes z-scores are given yet a statistician wants to know the raw scores instead. When z-scores, the mean, and the standard deviation are provided, you can plug what is known into the formula and solve for x. This becomes an algebra problem for which we need to isolate the unknown (x) on one side of the formula to solve, as shown in the example below. For this section we will presume we have population data and, thus, will use population symbols for the formula.

Formula and Steps	Explanation
Given: μ = 16.00 σ = 4.00 z_1 = 2.00	Start by identifying what is known and the formula needed.

Formula and Steps	**Explanation**
Find X_1: $z=\dfrac{x-\mu}{\sigma}$	
$2.00=\dfrac{x-16.00}{4.00}$	Plug in what is known. The remaining unknown is x.You can see that the value being solved for (which is x) is not isolated on one side of the equation. Thus, steps need to be used, one at a time, to isolate the unknown.
$2.00(4)=\dfrac{x-16.00}{4.00}(4.00)$	Remove divide by 4 by doing its opposite to both sides. This means we must multiple each side of the equation by 4.
$8.00=x-16.00$	Things are looking simpler but the x is still not isolated so an additional step is needed
$8.00+16.00=x-16.00+16.00$	Remove subtract 16 by doing its opposite to both sides. This means we must add 16 to each side of the equation.
$24.00=x$	We have solved for x and now know it was 24.00
$x=24.00$	Optional: Some people prefer to reorder the result so it reads from left to right the same way we would typically say it "$x\ equals$ 24.00"

We can follow this same process to convert any z-score to its corresponding raw score for a data set. Here are three examples of different raw scores being converted following the example from above:

Example 1	Example 2	Example 3
Given: μ= 16.00 σ = 4.00 z_1 = 21.00 Find X_1: $z=\dfrac{x-\mu}{\sigma}$ $2.00=\dfrac{x-16.00}{4.00}$ $2.00(4.00)=\dfrac{x-16.00}{4.00}(4.00)$ $8.00=x-16.00$ $8.00+16.00=x-16.00+16.00$ $24.00=x$	Given: μ= 16.00 σ = 4.00 z_1 = 8.80 Find X_1: $z=\dfrac{x-\mu}{\sigma}$ $0.40=\dfrac{x-86.00}{7.50}$ $0.40(7.50)=\dfrac{x-86.00}{7.50}(7.50)$ $3.00=x-86.00$ $3.00+86.00=x-86.00+86.00$ $89.00=x$	Given: μ= 16.00 σ = 4.00 z_1 = 16.40 Find X_1: $z=\dfrac{x-\mu}{\sigma}$ $-1.50=\dfrac{x-55.00}{1.62}$ $-1.50(1.62)=\dfrac{x-55.00}{1.62}(1.62)$ $-2.43=x-55.00$ $-2.43+55=x-55.00+55.00$ $52.57=x$

The Purpose of z-Scores

Raw scores are converted to z-scores and used to create something known as a standard distribution. Standardizing a distribution refers to the act of transforming from a normal curve with raw scores to an equivalent normal curve with z-scores. Thus, a standard distribution is a normal curve with z-score locations as the anchors for the x-axis in place of their raw score counterparts. Standardizing in this way has several practical applications.

First, it makes it easy to immediately deduce whether the raw score was greater or less than the mean or other measures of central tendency including the median and the mode. By simply checking whether the z-score is positive or negative you can deduce whether the raw score was greater than or less than the mean, respectively.

Second, knowing the z-score makes it easy to tell how common and likely or rare and unlikely a score is. The closer the z-score is to 0, the more common it and its corresponding raw score are. The farther the z-scores is from 0, regardless of whether it is positive or negative, the rarer it and its corresponding raw score are.

Third, and relatedly, z-scores help us identify outliers. Outliers are extreme, rare scores which would be found in the tails of the graph. Outliers have large z-scores indicating they are rare and far from most other scores in a normal distribution. However, the boundaries for which z scores are considered outliers are a bit flexible and depend upon aspects of the variable and data such as sample size and importance of skew (which is caused by outliers). Because there is flexibility, it is useful to review a few z-scores and why they are or are not generally used as outlier boundaries.

Let's consider the boundaries for outliers in a normal distribution. Approximately 68.26% of raw scores are within one standard deviation of the mean and are not considered outliers in most circumstances. This means z-scores between -1.00 and 1.00 are not considered outliers. However, z-scores of ± 1.96 are sometimes used to identify outliers. This is because approximately 95.00% of raw scores are within 1.96 standard deviations of the mean; sometimes, therefore, z-scores beyond this range are considered outliers because they represent the rarest 5% of scores. Thus, when ± 1.96 is used as the cutoff or boundary for identifying rare scores, z-scores outside this range (such as z-score of -2.00 or 2.00) are considered outliers. The use of this boundary can be more useful when a sample size is small. However, you may also see other cutoffs recommended.

Two other cutoffs are also often considered: z-scores of ± 2.58 and z-scores of ± 3.00. Though there is some gray area about what is considered an outlier, z-scores at or beyond -2.58 and 2.58 are often considered outliers. This is because approximately 99.00% of scores are within ± 2.58 standard deviations of the mean (meaning they would be between z-scores of -2.58 and 2.58). When this boundary is being used, the statistician is defining the rarest 1% of scores as outliers.

However, others prefer using ± 3.00. This would mean that the 0.26% of scores, which are found outside the boundaries of -3.00 and 3.00, would be considered outliers and the other 99.74% of scores within the boundaries of -3.00 and 3.00 would not be considered outliers. Though this is a bit of a strange cutoff (because it defines the rarest 0.26% as outliers), it is sometimes used because of its ease. When a normal curve is sketched, the x-axis is often drawn showing z-score locations starting from -3.00 and ending at 3.00. Thus, when the normal curve is sketched, these boundaries are used because nearly all scores fall between them and it is simpler to show integers rather than decimal numbers such as 1.96 or 2.58. Thus, ± 3.00 is sometimes used as a simple rule of thumb for identifying outliers visually. Nevertheless, for some statisticians ± 2.58 is preferred over 3.00 when working with larger samples because it is based more on mathematical logic than convenience.

Fourth, z-scores allow for comparison of scores within a group/sample. By knowing the z-scores of two individuals in a group, it is easy to identify which is rarer (when data are normally distributed). Suppose the mean is 88.00 and the *SD* is 6.60. You have two raw scores of 89.65 and 84.70. Suppose you want to know which score is rarer. This is hard to quickly deduce from the information provided (but it is possible). However, suppose that instead of the two raw scores you have their z-scores which are 0.25 and -0.50, respectively. You don't need to look at the mean, the *SD*, or even the raw scores themselves to know which is more common (i.e. the z-score of 0.25) and which is rarer (i.e. the z-score of -0.50). Thus, using z-scores makes it easy to quickly compare scores within a group or sample.

The fifth reason z-scores get used is a bit more complex but is especially useful: z-scores allow us to compare scores of different quantitative variables to each other. For example, we may want to compare a student's score on an English exam to their score on a psychology exam to see if their relative performance on each was approximately equivalent. Relative here refers to how well they performed compared to others on each exam and on each exam compared to the other exam. Making a direct comparison of raw scores for each exam is limited but useful comparisons of standardized scores (z-scores) on each exam are possible. Suppose the student's score on the English exam was 80 points and their score on the psychology exam was 65 points. We may assume that the English score indicates the higher relative skill simply because is the larger number. However, we have no indication here as to how many points were possible on each exam nor how well others did on each exam. If the English exam was out of 200 points and the psychology exam was out of 100 points, then the grade as a percentage would actually be higher for the psychology exam (65%) than for the English exam (40%). This is a way of comparing data many are accustomed to but it cannot tell us how well the student did relative to others. For that, a standardized score, such as a z-score, is needed.

Using a z-score for each exam allows a statistician to identify on which exam, if either, a student had better performance relative to others. Let's take the example of the student with scores of 80 and 65 on an English exam and a psychology exam, respectively. Suppose the mean score on the English exam was 70 with a standard deviation of 10 and that the mean score of the psychology exam was 50 with a standard deviation of 5. Presume the means and *SD*s represent the population of students. We can find the student's z-score for each exam as follows:

Psychology Exam	Psychology Exam	Comparison

Psychology Exam	Psychology Exam	Comparison
Given: μ= 70.00 σ = 10.00 x_1 = 80.00 Find X_1: $z = \frac{x - \mu}{\sigma}$ $z = \frac{80.00 - 70.00}{10.00}$ $z = \frac{10.00}{10.00}$ $z = 1.00$	Given: μ= 50.00 σ = 5.00 x_1 = 65.00 Find X_1: $z = \frac{x - \mu}{\sigma}$ $z = \frac{65.00 - 50.00}{10.00}$ $z = \frac{15.00}{5.00}$ $z = 3.00$	The student had a higher relative score on the psychology exam than on the English exam.

Reading Review 5.3

Try to answer the following questions using the graph:

1. What are the symbols in the z-score formula and what does each stand for?
2. What are the steps to computing a z-score from a given data set?
3. What are five uses for z-scores?

5.4: Finding Probabilities

Estimating Sampling Probabilities with the Normal Distribution

As we have just reviewed, the areas under the curve can be used to represent and calculate the proportion of scores in those areas. This also represents the probability that any given score is in that location. Let's consider this in relation to random sampling. When we randomly sample a score, it means each individual score has an equal chance of being selected. This would mean scores that occur more frequently and which are closer to the peak of the curve are more likely to be randomly selected because there are more of them.

Finding Probabilities using z-Scores

Once a z-score is known, it can be used to calculate probabilities within the normal distribution. This is because the percentages corresponding to areas under the curve represent the probabilities of scores at the corresponding sections of the curve. To say it another way, the percentage of an area in the normal curve is equal to the probability of scores existing and being randomly selected from that area. For example, half of the scores in the normal distribution are expected to be below the mean while the other half are expected to be above it because the mean is the center of the symmetrical normal curve. This also means that the probability that any score selected at random is below the mean is 50% and that the probability that any score selected at random is above the mean is also 50%.

However, not all areas of the curve are as simple to calculate as the two halves. When finding the areas and their corresponding probabilities under the curve, this three step process is useful: 1. Sketch, 2. Shade, and 3. Solve. Start by sketching the normal curve to provide a visual representation of the problem. Next, identify the z-score or z-scores needed and create a vertical line extending up from the z-score(s) location(s) on the x-axis. Shade the area you are solving for. Take note of whether the shaded area includes or extends to the mean and/or includes or extends to one of the tails. 3. Use the z-distribution table to find the percent represented in the shaded region of the sketch. This percentage is equal to the probability of any random score being selected from that area. This three step process can be used to find probabilities below, above, and between z-scores and to compare those probabilities.

Reading the z-Distribution Table

The z-score table, also known as a unit normal table, is located in Appendix C. Here is an excerpt of the z-score table:

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
0.00	0.5000	0.5000	0.0000
0.01	0.5040	0.4960	0.0040
0.02	0.5080	0.4920	0.0080
0.03	0.5120	0.4880	0.0120
0.04	0.5160	0.4840	0.0160
0.05	0.5199	0.4801	0.0199
0.06	0.5239	0.4761	0.0239
0.07	0.5279	0.4721	0.0279
0.08	0.5319	0.4681	0.0319
0.09	0.5359	0.4641	0.0359
0.10	0.5398	0.4602	0.0398

The z-distribution table shows the areas within the normal curve that are created when a z score is used to divide the graph into segments. Each z-score appears in the first column and the portions it divides the graph into are provided in columns 2 through 4 in decimal form. To convert from decimals to percentages, the proportion must be multiplied by 100. Note that the graph is symmetrical so the absolute values of z-scores are used in column 1 of the z distribution table which allows the table to be used with positive or negative z-scores. However, we must be clear on what each column represents before we can use the table to find proportions of the normal curve:

- Column 2 represents the area from the z-score to its furthest tail. Thus, if a z-score is positive, this is the area from that z-score all the way through the negative tail. If a z score is negative, this is the area from that z-score all the way through the positive tail.
- Column 3 represents the area from a z-score to its nearest tail. For positive z-scores this is the area from the score through the positive tail and for negative z-scores this is the area from the score through the negative tail.
- Column 4 represents the area from a given z-score to the mean (which is at the very center of the graph). For positive z-scores this is the area that goes toward the left from the score until it reaches the center of the curve. For negative z-scores this is the area that goes toward the right from the score until it reaches the center of the curve.

Let's consider a z-score of 0 (or 0.00 when shown to the hundredths place). We know that this is the center of the curve so the area from it to either tail is 50.00%. The tails are equidistant from the mean so columns 2 (area to the furthest tail) and 3 (area to the closest tail) are both .5000 (which corresponds to 50.00%). There are no scores between a z-score of 0.00 and the mean and, thus, column 4 shows that .0000 (or 0.00%) of score are between a z-score of 0.00 and the mean.

We have briefly reviewed how to read and understand the z-distribution table here but this concept can be complex and used in many different ways. Therefore, we will review seven different ways proportions of the normal curve can be divided and estimated using the z distribution table in detail in the following sections.

Finding Probabilities Below z-Scores

Using Negative z-Scores. The area below a negative z-score will always be less than 50.00%. This is because the whole area below the mean is only 50.00% so any portion of that area must be less than 50.00%. Let's take a look at an example using the three step process. Suppose you are asked to find the probability that a score below z = -1.50 will be randomly selected. First, sketch the normal curve. Second, draw a vertical line over the location of z = -1.50. Shade the area to the left of that line because lower scores are to the left on the x-axis. This shaded region represents the proportion of the scores that are below a z-score of -1.50 (See Figure 3a). Notice that the shaded area is a small portion of the graph. This is a visual indication that the percentage corresponding to the area will also be small. The shaded area goes from the z-score to the nearest tail so this is the region we need to look up in the z-score table.

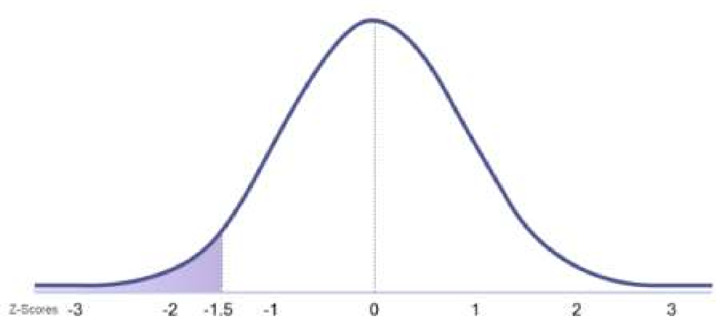


Figure 3a.

We can use the z-distribution table to find the area that corresponds to the area from z = -1.50 to the tail. Remember, the z-distribution table shows the absolute value of each z-score because the normal curve is symmetrical. Therefore, the area from -1.50 to its nearest tail (which is the tail on the left) and from 1.50 to its nearest tail (which is the tail on the right) are the same. Because of this, we can look up the area from 1.50 to the nearest tail in the table and use it to represent the area from -1.50 to its nearest tail. This area is 0.0668 in decimal form or 6.68% when reported as a percentage. Therefore, the probability of a score being randomly selected from below the z-score of -1.50 is 6.68% (See Figure 3b).

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
1.49	0.9319	0.0681	0.4319
1.50	0.9332	**0.0668**	0.4332
1.51	0.9345	0.0655	0.4345
1.52	0.9357	0.0643	0.4357
1.53	0.9370	0.0630	0.4370
1.49	0.9319	0.0681	0.4319
0.55	0.7088	0.2912	0.2088

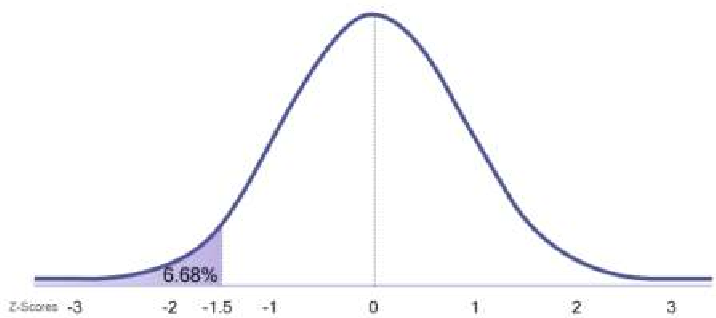


Figure 3a with Proportion Shown.

Using Positive z-Scores. The area below a positive z-score will always be more than 50.00%. This is because the area will start to the right of the mean and extend all the way to the left, subsuming the entire lower half (50.00%) of the graph. Let's take a look at an example using the three step process. Suppose you are asked to find the probability that a score below z = 1.50 will be randomly selected. First, sketch the normal curve. Second, draw a vertical line over the location of z = 1.50. Shade the area to the left of that line because lower scores are to the left on the x-axis; the shaded region goes to and past the mean all the way into the negative (lower) tail of the graph. This shaded region represents the proportion of the scores that are below a z score of 1.50 (See Figure 4). Notice that the shaded area is a large portion of the graph. This is a visual indication that the percentage corresponding to the area will also be large.

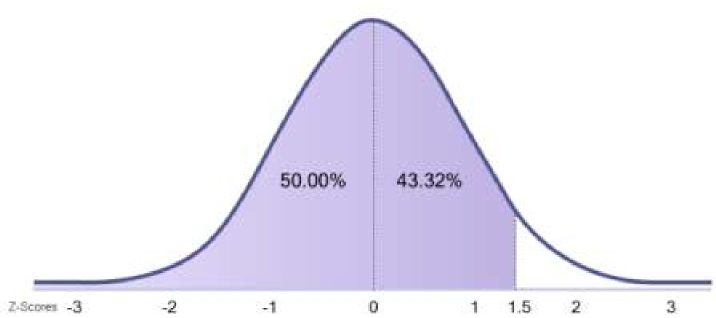


Figure 4.

We already know the area below the mean is 50.00% but need to find the area from z = 1.50 to the mean to get the total area shaded. We can use the z-distribution table to find the area that corresponds to the area from z = 1.50 to the mean. This area is 0.4332 in decimal form or 43.32% when reported as a percentage. We can add this to the 50.00% that is below the mean to get the following: 43.32% + 50.00% = 93.32%. Alternatively, we could look up column 2 of the z-distribution table as it shows the area from a z-score to its furthest mean (which is what we need here). As expected, we see this is also 93.32%. Therefore, the probability of a score being randomly selected from below the z-score of 1.50 is 93.32%.

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
1.49	0.9319	0.0681	0.4319
1.50	**0.9332**	0.0668	0.4332
1.51	0.9345	0.0655	0.4345
1.52	0.9357	0.0643	0.4357
1.53	0.9370	0.0630	0.4370
1.49	0.9319	0.0681	0.4319
0.55	0.7088	0.2912	0.2088

Finding Probabilities Above z-Scores

Using Positive z-Scores. The area above a positive z-score will always be less than 50.00%. This is because the whole area above the mean is only 50.00% so any portion of that area must be less than 50.00%. Notice that this the same logic applied to areas below a negative z-score. Suppose you are asked to find the probability that a score above z = 1.50 will be randomly selected. First, sketch the normal curve. Second, draw a vertical line over the location of z = 1.50. Shade the area to the right of that line because higher scores are to the right on the x-axis. This shaded region represents the proportion of the scores that are above a z-score of 1.50 (See Figure 5). Notice that the shaded area is a small portion of the graph. This is a visual indication that the percentage corresponding to the area will also be small. The area shaded goes from the z-score to the tail so this is the region we need to look up in the z-score table.

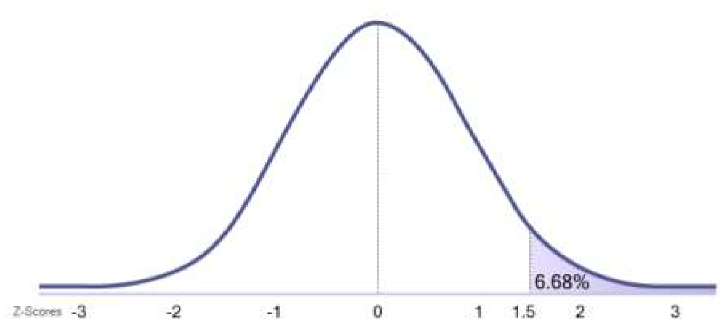


Figure 5.

We can use the z-distribution table to find the area that corresponds to the area from z = 1.50 to the nearest (right or upper) tail. This area is 0.0668 in decimal form or 6.68% when reported as a percentage. Therefore, the probability of a score being randomly selected from above the z-score of 1.50 is 6.68%. This is the same as the probability of an area below a z-score of -1.50 because the graph is symmetrical and these represent areas which mirror each other in the normal curve.

Using Negative z-Scores. The area above a negative z-score will always be more than 50.00%. This is because the area will start to the right of the mean and extend all the way to the left, subsuming the entire lower half (50.00%) of the graph. Notice that this is the same logic applied to areas below positive z-scores. Let's take a look at an example using the three step process. Suppose you are asked to find the probability that a score above z = -1.50 will be randomly selected. First, sketch the normal curve. Second, draw a vertical line over the location of z = -1.50. Shade the area to the right of that line because higher scores are to the right on the x axis; the shaded region goes to and past the mean all the way into the positive (upper) tail of the graph. This shaded region represents the proportion of the scores that are above a z-score of -1.50 (See Figure 6). Notice that the shaded area is a large portion of the graph. This is a visual indication that the percentage corresponding to the area will also be large.

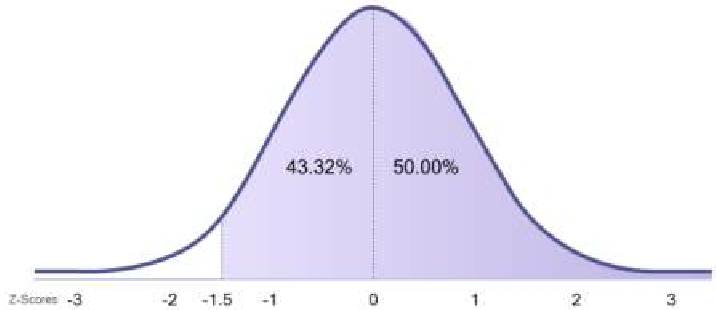


Figure 6.

We have two options for finding the shaded region. We already know the area above the mean is 50.00% but need to find the area from z = -1.50 to the mean to get the total area shaded. We can use the z-distribution table to find the area that corresponds to the area from z = -1.50 to the mean (Remember, the z-distribution table shows absolute values of z so we still need the row that corresponds to z = 1.50). This area is 0.4332 in decimal form or 43.32% when reported as a percentage. We can add this to the 50.00% that is above the mean to get the following: 43.32% + 50.00% = 93.32% Therefore, the probability of a score being randomly selected from above the z-score of -1.50 is 93.32%. Alternatively, we can use column 2 of the table which shows the area from the z-score to the furthest tail. Alternatively, we could look up column 2 of the z-distribution table as it shows the area from a z-score to its furthest mean (which is what we need here). As expected, we see this is also 93.32%. The two methods provide the same result so either can be used. Notice that the area above z -1.50 is the same as the probability of an area below a z-score of 1.50 because the graph is symmetrical and these represent areas which mirror each other in the normal curve.

Finding Probabilities Between z-Scores

Finding area between two z-scores is a little more complicated and depends on the direction of the two scores in relation to each other. In order to find the area or probability of a score being randomly selected from between two z-scores, you must first identify whether the two z-scores have opposing signs (meaning one is positive and one is negative) or whether both have the same sign (meaning either both are positive or both are negative). Then, we follow the three step process of sketching, shading, and solving.

Using a Negative and a Positive z-Score. Suppose we want to find the area (or probability of randomly selecting a score) between z = -0.50 and z = 0.75. First, sketch the normal curve. Second, draw a vertical line over the location of z = -0.50 and another over the location of z = 0.75. Shade the area between the two vertical lines. This shaded region represents the proportion of the scores that are between the two z-scores (see Figure 7).

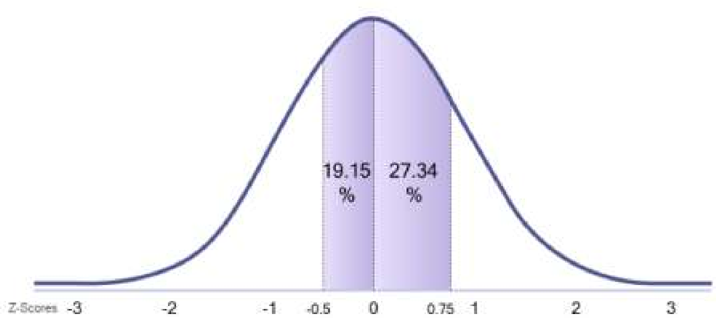


Figure 7.

Keep in mind that the z-distribution table does not tell us area between two z-scores; instead it tells us the areas from each z-score to the mean and to the nearest tail. Notice, however, that the mean is between the two z-scores and is, thus, within the shaded region. This means that the area we are solving for is actually the area from each of the z-scores to the mean added together. Therefore, for step 3 of the process, we must find the area from each z-score to the mean and then add them to get the total for the shaded area. The area from z = -0.50 to the mean is 0.1915 in decimal form or 19.15% when reported as a percentage (see Appendix C for the full z-distribution tables). The area from z = 0.75 to the mean is 0.2734 in decimal form or 27.34% when reported as a percentage. We can add these two areas together to get the total for the shaded area as follows: 19.15% + 27.34% = 46.49% Therefore, the probability of a score being randomly selected from between z-scores of -0.50 and 0.75 is 46.49%.

Using Two Positive z-Scores

To find the area between two positive z-scores, we must find an area that goes neither to the mean nor to the tail so it is a bit unique. Let's illustrate this with an example. Suppose we want to find the area (or probability of randomly selecting a score) between z = 0.50 and z = 0.75. Start our three step by process by first sketching the normal curve. Second, draw a vertical line over the location of z = 0.50 and another over the location of z = 0.75. Shade the area between the two vertical lines. This shaded region represents the proportion of the scores that are between the two z-scores (see Figure 8a).

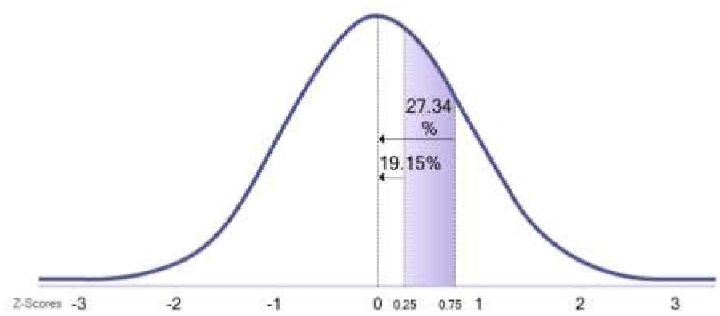


Figure 8a.

The z-distribution table will not show us this exact region so we have to get creative to solve for the area in step 3 of the process. Start by finding the area from z = 0.75 to the mean; this area is 0.2734 in decimal form or 27.34% when reported as a percentage. However, we don't want this total area. Instead, we only want the portion of this area that ends at z = 0.50. Therefore, we want to remove (subtract) the area that goes past 0.50 and to the mean so we will be left with only the shaded region. To do this, we must next find the area from z = 0. 50 to the mean which is 0.1915 in decimal form or 19.15% when reported as a percentage. Finally, we must subtract the smaller area from z = 0.50 to the mean from the larger area of z = 0.75 to the mean to find the shaded area as follows: 27.34% - 19.15% = 8.19% Therefore, the probability of a score being randomly selected from between z-scores of 0.75 and 0.50 is 8.19% (see Figure 8b).

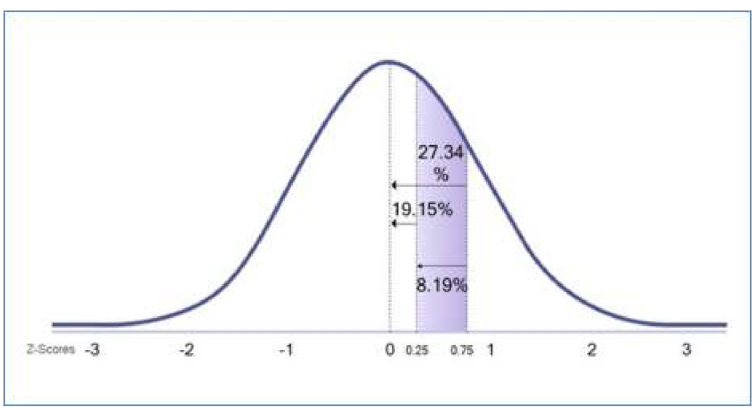
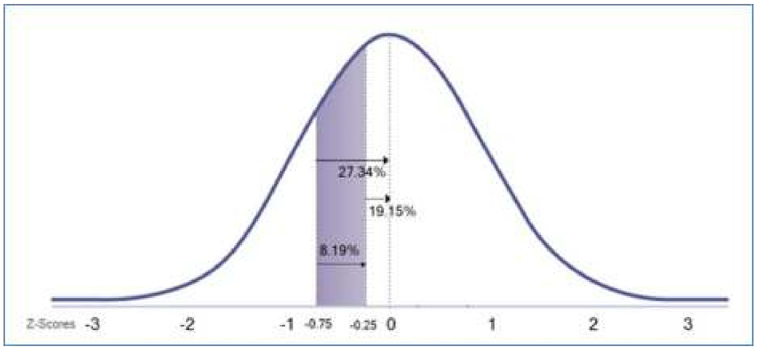


Figure 8b.

Using Two Negative z-Scores

Finding the area between two negative z-scores requires the same logic and process as was used to find the area between two positive z-scores. Let's illustrate this by using the negative version of the prior example. Suppose we want to find the area (or probability of randomly selecting a score) between z = -0.50 and z = -0.75. Start the three step by process by first sketching the normal curve. Second, draw a vertical line over the location of z = -0.50 and another over the location of z = -0.75. Shade the area between the two vertical lines. This shaded region represents the proportion of the scores that are between the two z-scores (see Figure 9).



Next, find the area from z = -0.75 to the mean; this area is 0.2734 in decimal form or 27.34% when reported as a percentage. However, just as we saw when looking for the area between two positive z-scores, we don’t want this total area. Instead, we only want the portion of this area that ends at z = -0.50. Therefore, we want to remove (subtract) the area that goes from - 0.50 to the mean so we will be left with only the shaded region. To do this, we must next find the area from z = -0. 50 to the mean which is 0.1915 in decimal form or 19.15% when reported as a percentage. Finally, we must subtract the smaller area of z = -0.50 to the mean from the larger area of z = -0.75 to the mean to find the shaded area as follows: 27.34% - 19.15% = 8.19% Therefore, the probability of a score being randomly selected from between z-scores of -0.75 and -0.50 is 8.19%. Notice that this is the same result we got in the prior section when using z = 0.50 and z = 0.75 because the symmetrical nature of the graph causes the left half of the graph to mirror the right half of the graph.

Comparing Probabilities using z-Scores

Comparisons of probabilities can be made using z-scores in a variety of scenarios. First, you can deduce which of two z-scores is more frequently occurring (or more probable) by determining which is closest to z = 0. For example, z = 0.50 is more probable than z = 1.00 because 0.50 is closer to 0 than is 1.00. Second, you can determine which raw scores are most probable by using their z-scores and following the logic in the prior example (i.e. whichever has a z-score closer to 0 is the more probable score). Third, you can compare probabilities of scores being selected from different areas under the curve. For example, we can determine whether there is a greater probability of selecting a z-score between z = 1.00 and z = 2.00 than selecting a z score between z = 2.00 and z = 3.00. It is tempting to focus on the fact that the distance from 1.00 and 2.00 on the x-axis is the same as the distance between 2.00 and 3.00 on the x-axis; this temptation can lead us to erroneously conclude that the probabilities of scores in these two ranges are equal. However, the normal curve slopes and probability is about area in these ranges, not simply distance on the x-axis. If we follow the steps for finding area between z scores, we will find that the area between z-scores of 1.00 and 2.00 is 13.59% and that area between z-scores of 2.00 and 3.00 is 2.15%. This means the probabilities of selecting scores in the aforementioned ranges are 13.59% and 2.15%, respectively. Thus, the probability of randomly selecting a z-score between 1.00 and 2.00 is about six times as high as the probability of randomly selecting a z-score between 2.00 and 3.00.

Finding Percentiles with z-Scores

z-scores can be used to determine percentiles. Percentiles indicate the proportion of scores or cases that are below a given score (see the sections on Percentile Rank Distributions in Chapter 2 to review how percentiles are calculated). For example, if we say someone’s IQ is in the 67% percentile, it means that their IQ is higher than 67% of the scores in the population and that their IQ is lower than 33% of the scores in the population. The total area to the left of a z-score in the normal curve is equivalent to that z score’s percentile. This is because when we solve for area below (left of) a score, we are identifying the proportion of scores lower than it; this is equivalent to the score’s percentile. Thus, there is no special computation or step needed to find percentiles using z-scores. Instead, we simply follow the process for finding area below a z-score. The difference is how we language (or report) the results. When solving for area below a z-score of 0.00, for example, we could report probability or percentile as follows:

Probability*:* The probability of randomly selecting a score below z = 0.00 is 50%

Percentile*:* A z-score of 0.00 is at the fiftieth percentile.

Though percentiles are often reported to the whole number for simplicity, it is possible to compute them with more specificity such as by rounding to the hundredths place, just as we have done with z-scores in this chapter.

Finding Raw Scores from Percentiles

Because of the relationships between z-score and percentiles, percentiles can be used to find raw scores if the mean and standard deviation are also known. Suppose you know that the mean IQ score is 100.00 with a standard deviation of 15.00 and are told an individual was in the 67% percentile for IQ. If you wanted to find their raw score, you would follow these steps:

1. Find the z-score that corresponds to the 67th percentile. Note: this is going to be a positive z-score because the percentile is over 50%.
2. Plug that z-score, the mean, and the standard deviation into the z-score formula.
3. Solve for X to find the raw score.

In the example above, we would find the following.

1. The z-score corresponding to the 67th percentile is 0.44 (as shown in the segment of the z-distribution table).

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
0.43	0.6664	0.3336	0.1664
0.44	**0.6700**	0.3300	0.1700
0.45	0.6736	0.3264	0.1736

2. Plug z = 0.44, μ = 100.00 and σ = 15.00 into the formula $z=\frac{x-\mu}{\sigma}$ as follows:

$$0.44=\frac{x-100.00}{15.00}$$

3. Solve for x as follows:

$$\begin{aligned}
&0.44=\frac{x-100.00}{15.00}\\
&0.44(15.00)=\frac{x-100.00}{15.00}(15.00)\\
&6.60=x-100.00\\
&6.60+100.00=x-100.00+100.00\\
&106.60=x
\end{aligned}$$

This result can be reported as follows:

Summary: An IQ score of 106.60 is in the 67th percentile.

5.5: Probability and Inferential Statistics

Probability is a tenet of statistics (see Chapter 1 for a review) and is, thus, a core component of both descriptive and inferential statistics. Probabilities and z-scores as they have been described so far within this chapter, have remained in the realm of descriptive statistics. Inferential statistics build on these concepts and use probabilities to determine what is likely true about populations based on data from samples. This is done through a variety of techniques constructed to address different kinds of data and hypotheses. What many of these techniques have in common is they:

1. Build from descriptive summaries of sample data,
2. Use those sample data and summaries to estimate patterns or differences for populations, and then
3. Calculate the probability that those population estimates are true based on the amount and strength of the data available from the sample.

In keeping with the first commonality noted above, we have already reviewed descriptive statistics starting from Chapter 2 and have begun working with probabilities in the present chapter (Chapter 5). In our next chapter (Chapter 6), we will deepen our understanding of the three tenets (variability, probability, and uncertainty) and their connections. Chapter 6 will also provide the knowledge base needed before we can use sample data to estimate population parameters (commonality 2 from above) and calculate the probability these estimates are true of populations (commonality 3 above). Therefore, following what we have just learned about z scores and probabilities, we can begin to shift from understanding descriptive statistics to exploring inferential statistics and will do just that starting in Chapter 6.

Reading Review 5.4

1. What is the probability of randomly selecting a raw score greater than $z = 1.00$?
2. What is the probability of randomly selecting a raw score lower than $z = 1.00$?
3. What is the probability of randomly selecting a raw score greater than $z = -0.85$?
4. What is the probability of randomly selecting a raw score lower than $z = -0.85$?
5. What is the probability of randomly selecting a raw score between than $z = -0.85$ and $z =1.00$?

CHAPTER OVERVIEW

6: The Foundations of Hypothesis Testing

6.1: Developing Hypotheses

Inferential statistics are used to test hypotheses. This is generally done by testing data from samples to learn what is likely true of populations. Hypothesis testing is central to behavioral sciences such as psychology. Before we can learn *how* to use inferential statistics to test hypotheses, we must first become familiar with two main things:

1. The types of, and details surrounding, hypotheses, and
2. The foundational concepts that connect and differentiate samples from populations.

Reviewing these two areas will provide the necessary foundation before we can embark on hypothesis testing.

Developing Hypotheses

Hypothesis testing is central to our work as researchers and statisticians. Two important skills to develop are the ability to generate testable hypotheses and the ability to correctly choose and use strategies to test those hypotheses. Researchers follow the scientific method by starting with making observations and reviewing existing knowledge. Therefore, researchers often have a general topic or question in mind and will then read existing, relevant research and theories to refine their question. When they have narrowed in, they will have a research question. As is overtly stated in the name, a **research question** is a question that a researcher or statistician intends to answer by analyzing data. When research is more exploratory in nature, the researcher may move forward solely with the research question to guide their work. However, scientific fields are aided by moving from research questions to hypotheses before planning for and collecting data. **Research hypotheses** are testable, expected answers to research questions that draw on larger theories, previous findings, and/or strong arguments. These forms of hypotheses can also be referred to as *alternative hypotheses* or simply as *hypotheses.* Designing studies, collecting data, and analyzing data depend upon the hypotheses that foment the research. Thus, inferential statistics and all the components that go along with it follow the formation and clear statement of a research hypothesis.

Stating a Research Hypothesis

Research hypotheses should be clear and specific, yet also succinct. A hypothesis should also be testable. If we state a hypothesis that is impossible to test, it forecloses any further investigation. To the contrary, a hypothesis should be what directs and demands investigation. In addition, a hypothesis should be directional, when possible. A **directional hypothesis** is one that includes information about the mathematical movement or difference that is expected; the direction tells which way you think the pattern(s) in the data will go. For example, if a researcher hypothesizes that teenagers and adults have *different* mean hours of sleep, their hypothesis is non-directional. This is because the hypothesis states a difference is anticipated

without specifying which group will have more or less sleep than the other. It could be that teenagers sleep more and adults sleep less or that teenagers sleep less and adults sleep more. In this example there are two opposing outcomes that could support the hypothesis. However, it is preferable to narrow the hypothesis down so that only one, specific outcome would support the hypothesis. If a researcher hypothesized that teenagers would sleep *more* hours than adults, their hypothesis would be directional. There is only one pattern in the data that could support this hypothesis: that the teens had comparatively *more* hours of sleep and adults had comparatively less sleep. Hypotheses such as these should be used whenever there is good reason to narrow in to one direction.

Whenever a hypothesis is stated, a hypothesis that counters it is also simultaneously being proposed which is known as a null hypothesis. A **null hypothesis** is a statement about a population that counters a research hypothesis and which is presumed to be true until there is sufficient evidence to refute or reject it. Because the null hypothesis is presumed to be true until there is sufficient evidence to support the research hypothesis and simultaneously reject the null hypothesis, research hypotheses are often called *alternative* hypotheses to reiterate that they are proposed alternatives to what is otherwise presumed to be true (which is the null hypothesis).

The assumption that the null hypothesis is true may seem counterintuitive at first as hypotheses are often grounded in theories and/or prior research, but there is an important reason for this: it requires that a researcher tests their presumptions (which are stated as their hypotheses) before they can present those presumptions as possible truths. This is what moves a field from being purely philosophical to empirical and, ultimately, scientific. Recall that a hypothesis should be testable. When a hypothesis is testable it means it is also falsifiable. **Falsifiable** means that something can be disproven or shown to be false if indeed it is not true. The onus is on the researcher to develop testable (and, thus, falsifiable) hypotheses and to test them before putting those hypotheses forth as possible facts. This process serves as an important filter between untested ideas (which can be stated as hypotheses) and supported contentions (hypotheses which have been supported by evidence).

Symbols

Abbreviations and symbols are often used to state the null and alternative (i.e. research) hypotheses. Population symbols are used for the null hypothesis. An important property of alternative hypotheses is that they describe predicted values of population parameters (not sample statistics). The alternative hypothesis, therefore, can be written with the population symbols to indicate that the researcher expects the hypothesis to be true beyond the sample. However, the alternative hypothesis can also be written with sample symbols to reiterate that the data used to test it are from samples and can only be used to estimate the population. This is because one of the principles of hypothesis testing is that we determine the sample statistics in order to infer the population parameters. Thus, sample symbols are sometimes used for alternative hypotheses to emphasize the distinction between what is estimated through the use of sample data from what is true about populations.

The name "alternative hypothesis" can be abbreviated with the symbols H_a where the H stands for the word "hypothesis" and the subscript *a* abbreviates the word "alternative." However, it is common for researchers to test more than one hypothesis in a single study. When doing so, it would be confusing to list several hypotheses with the same symbols. Therefore, it is also appropriate to enumerate the hypotheses in the order in which they will be tested using consecutive numbers in place of *a* in the subscript. Thus, H_1 would be used as the abbreviation for the first hypothesis, H_2 for the second hypothesis, H_3 for the third hypothesis, and so on until each alternative hypothesis has been specified.

The null hypothesis is abbreviated with the symbols H_0 because zero is synonymous with the word null. When a researcher states an alternative hypothesis, the null hypothesis is presumed and typically does not need to be overtly stated. Thus, null hypotheses are often all referred to simply as H_0 without additional enumeration when they are written because they will be presented alongside their corresponding alternative hypothesis, if presented at all.

However, sometimes a researcher will overtly state both the alternative hypothesis and its corresponding null hypothesis (especially when they are taking a class and first learning how alternative and null hypotheses work). The researcher also makes the decision about whether the alternative hypothesis will be directional or non-directional based on information obtained from a review of prior research and theories. The alternative and null hypotheses for a given research question are mutually exclusive. Mutually exclusive means that if one of these two forms of hypothesis is true, the other one must be false. For example, only one of the following hypotheses in each pair can be true:

Pair 1

- **Directional, alternative hypothesis:** Last-minute studying will increase students' understanding of inferential statistics.
- **Null hypothesis:** Last-minute studying will not increase students' understanding of inferential statistics.

Pair 2

- **Non-directional, alternative hypothesis:** Last-minute studying will have an effect on students' understanding of inferential statistics.
- **Null hypothesis:** Last-minute studying will have no effect on students' understanding of inferential statistics.

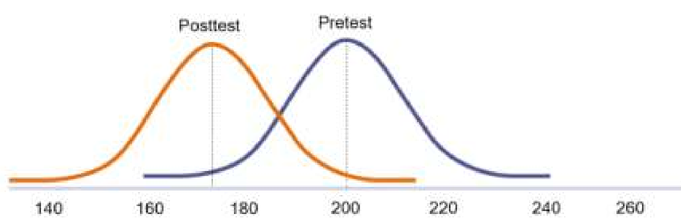
Data can only support one statement in each pair and refute the other because each pair of hypotheses (alternative and null) are mutually exclusive.

6.2: One-Tailed vs. Two-Tailed Tests

Tests of many hypotheses can be categorized as one-tailed or two-tailed dependent upon whether the hypotheses themselves are directional or non-directional. Generally, directional hypotheses require one-tailed tests and non-directional hypotheses require two-tailed tests.

The names one-tailed and two-tailed refer to whether one or both tail regions of the normal curve are being considered in the stated hypothesis. Think of it this way: if you start at the center of the normal curve there are two directions you can look to see if there are patterns or groups of data elsewhere. There could be groups of data to the left (lower tail) or to the right (upper tail) of the center. If your hypothesis is directional and states that data are expected to be lower (or lesser), it means you only need to look in one direction which is to the left (lower tail) to see if the data are there. If your hypothesis is directional and states that a group of data are expected to be higher (or greater), it means you only need to look in one direction which is to the right (upper tail) to see if the data are there. However, if a hypothesis is non-directional it means that a group of data is expected to be somewhere other than the middle; it could be to the left or it could be to the right of the center meaning that you must look both directions to check each of the two tails for those data. Thus, when a direction is given to a test using the normal distribution and only one-tail needs to be checked, the test is called one-tailed and when it is non-directional such that both tails need to be checked, it is called two-tailed.

Let's review an example using a directional hypothesis. Suppose it is hypothesized that cholesterol levels will be *lower* after eating oatmeal daily for six weeks compared to before. In this example the population parameter of interest is the mean cholesterol level. Data would be collected about cholesterol to start and graphed as a normal curve. These are known as *pretest* data. Then data would be collected about cholesterol again at the end of the six weeks; these are known as *posttest* data and would be graphed as a normal curve to see if there has been any change. The mean which is the center of the posttest normal curve is expected to have shifted to the left compared to the pretest normal curve when both are plotted over the same x-axis. Therefore the alternative hypothesis indicates that the posttest mean is expected to be to the left of the pretest mean. The null hypothesis counters this and, thus, expects that either the posttest mean will be the same as the pretest mean or that the posttest mean will actually be higher (to the right) of the pretest mean.



Directional, alternative hypothesis: Cholesterol levels will be lower after (post) eating oatmeal daily for six weeks compared to before (pre).

$$H_a : \mu_{\text{post}} < \mu_{\text{pre}}$$

Null hypothesis: Cholesterol levels will not be lower after (post) eating oatmeal daily for six weeks compared to before (pre).

$$H_0 : \mu_{\text{post}} \geq \mu_{\text{pre}}$$

Reading Review 6.1

1. The three names which can be used for an alternative hypothesis are alternative hypothesis, __________________________, or simply ______________.
2. What is the difference between an alternative hypothesis and a null hypothesis?
3. Which hypothesis is presumed to be true until refuted by evidence?
4. What is the difference between a directional and a non-directional hypothesis?
5. For which hypotheses would a one-tailed test be used instead of a two-tailed test?

6.3: Sampling Distributions

In previous chapters we have focused on how to summarize data from samples by looking at one sample at a time. For example we computed means, standard deviations, and even z-scores to summarize a sample's distribution (through the mean and standard deviations) and to estimate the expected locations and probabilities of individual raw scores within a distribution (through transforming to and using z-scores). The practices connects out to broader assumptions about populations and how data from samples may represent those populations.

In statistics, samples are used to estimate populations based on an assumption of how the pattern of data from many samples can represent populations. To understand this, we must first consider a concept known as a sampling distribution and how this is connected to an idea known as the Central Limit Theorem.

Let's zoom out from raw scores as the unit of analysis to sample means. Suppose that instead of considering raw scores for many cases as a way to estimate a variable, we used samples as the cases and represented them with their means. Another way to say this is that the means from many samples would be used in place of raw scores for computations. The mean of those means and the standard deviation of those means could then be used to construct things the normal curve and to estimate z-scores and probabilities to represent the population of interest. Patterns and variation in data for a variable (which would be represented in the normal curve) are referred to as *distributions.* When we use data from one sample, we refer to the patterns and dispersion as the *distribution of sample data.* When we zoom out and use means in place of raw scores, we refer to the patterns and variation as a *sampling distribution.*

Here is the boiled down explanation of what is assumed about sampling distributions: Samples are drawn from populations and are used to estimate population means. If this were to be done with replacement (meaning the full population is being sampled from each time) and a sufficient number of random samples of the population are taken, it would be called the sampling distribution. Thus, a **sampling distribution** is like a data set but with sample means in place of individual raw scores. When these samples are drawn randomly and with replacement, most of their means are expected to be fairly close to the true population mean and few are expected to be far from the true population mean. In fact, if samples are gathered from a population over and over again, the distribution of those sample means is expected to form a normal distribution the same way individual raw scores are expected to within any given sample. In this way, the distribution of many sample means is essentially expected to recreate the actual distribution of scores in the population if the population data are normal. However, even if the data in the population are skewed or are randomly generated, the sampling distribution is expected to be normal. Thus, **Central Limit Theorem** states that the sampling distribution will tend to be normal in many situations if a sufficient number of samples of sufficient size are drawn randomly and with replacement. Notice that, once again, we are stating what tends to, or is likely, to happen but not what is guaranteed. Thus, Central Limit Theorem states what is probable or tends to be true rather than what is absolute or guaranteed to occur every time.

Standard Error

Sampling distributions have means and something known as a standard error (*SE*) which together function like the mean and standard deviation (*SD*) do for individual samples. The distinction between the means is that the sampling distribution uses a mean of sample means whereas a sample uses a mean of raw scores. Both the *SE* and *SD* describe how far things tended to deviate from a mean. Thus, each is estimating how much error there is when using its respective mean to estimate something. However, these differ in what they are trying to estimate and, thus, what the error is in regards to. Specifically, *SD* focuses on how far raw scores in a sample tended to be from their mean while *SE* is used to estimate how far the mean of a sampling distribution is from the population mean it is being used to estimate.

Standard error is calculated by adjusting a *SD* by the square root of sample size using this formula:

$$SE = \frac{s}{\sqrt{n}}$$

Let's take a look at how to calculate standard error with an example.

✔ Calculating Standard Error (SE)

Given:

s = 9.3859063 …

n = 22

Find SE

Solution

$$SE = \frac{s}{\sqrt{n}}$$
$$SE = \frac{9.3859\ldots}{\sqrt{22}}$$
$$SE = \frac{9.3859\ldots}{4.6904\ldots}$$
$$SE \approx 2.0010\ldots$$

The SE can be rounded to the hundredths place and summarized as: SE = 2.00.

Sampling Error

Sampling error refers to the amount of error (or difference) between a sample statistic and the parameter in the population the sample represents. Each time a sample is drawn and a statistic such as the mean is computed, it may be discrepant from the population mean. However, as the sample size increases, more of the population is being represented by the sample and, thus, the less sampling error is expected. Standard error is a specific calculation that is used to estimate the sampling error between any one sample mean and its population mean. Therefore, sampling error refers to the broader idea that sample statistics (such as the mean) are rarely perfect representations of population parameters while standard error estimates the amount a sample mean deviates from the population mean it is being used to represent. Because population means are rarely known, standard errors are a necessary consideration when using sample means to estimate unknown population parameters.

Though there is much more that can be said about sampling distributions, Central Limit Theorem, standard errors, and sampling error, this boiled down review focused on the attributes and scenarios most applicable to the level of understanding required to move forward to probability and hypothesis testing.

Standard errors (SE)

Standard errors (SE) summarizes the average deviation between sample means ($\bar{x}$) and the population means (μ) they are being used to estimate.

SE is calculated by dividing standard deviation (s) by the square root of the sample size (n).

The symbols $\sigma_{\bar{x}}$ or σ_M are sometimes used in place of SE to emphasize that the standard error of sampling means is an adjusted version of standard deviation.

Reading Review 6.2

1. What is a sampling distribution?
2. What does Central Limit Theorem state should occur when a sufficient number of samples of sufficient size are drawn randomly and with replacement?
3. Which statistic is calculated to estimate sampling error?

6.4: Hypothesis Testing

Probability and Hypothesis Testing

Hypothesis testing is a process by which data are analyzed and conclusions are drawn about whether the results support or refute the hypothesis. This process allows statisticians to determine the likelihood that their results are not due to chance and, instead, likely represent truths about populations that are in keeping with their hypotheses. Note that the tenet of probability (introduced in Chapter 1) is a component of this process. When a hypothesis is tested, steps are followed and calculations are performed to assess the *probability* (likelihood) that a hypothesis is true based on sample data. Thus, the terms supported and refuted are used instead of the words proven and disproven, respectively.

Hypotheses come in a variety of forms, each of which requires different statistical methods of analysis. Some hypotheses, like the one about oatmeal and cholesterol from earlier in this chapter, state that a treatment condition (or intervention) will cause a difference in a measurable outcome variable. To review, that hypothesis written in sentence and symbol formats is:

Cholesterol levels will be lower after (post) eating oatmeal daily for six weeks compared to before (pre).

$$H_a : \mu_{\text{post}} < \mu_{\text{pre}}$$

In this example daily oatmeal consumption is the treatment condition and cholesterol level is the outcome being measured. We can see from the symbol format that data will need to be collected from a sample both before the treatment in order to compute a pretest mean and again from the same sample after the treatment in order to compute a posttest mean. The means can then be compared to see if, as is expected based on the hypothesis, the mean cholesterol level for the sample is lower at posttest than it was at pretest.

It is tempting and seems logical to simply conclude that if the posttest mean is even slightly lower than the pretest mean, the hypothesis is supported. However, before we can draw this conclusion we need to assure that our results are strong enough to conclude that the difference is unlikely to simply be due to chance and that they, instead, likely reflect a real difference in the means. This is because sample means are estimates and are expected to have some sampling error; sample means are not expected to be perfect representations of population parameters under the same conditions. Slight differences in means could simply be due to sampling error. Thus, estimates of error (such as the standard deviation or standard error) must also be considered in order to determine how likely it is that the difference in the pretest and posttest sample means represent an actual difference that would be observed in population parameters under the same conditions. For this reason, estimates of error are an important part of statistical power and determining significance.

Statistical Power and Significance

Statistical power refers to how likely it is that sample data will support the hypothesis. Think of statistical power as the ability to detect that an alternative hypothesis is true if, in fact, it is true. Power is generally increased or decreased by three factors:

1. the size of the sample,
2. the size of the change, difference, or pattern observed in the sample data, and
3. the size of the error in the relevant estimates of the changes, differences, or patterns observed.

First, if the hypothesis is true, data to support it are more likely when the sample is larger than when it is smaller. Therefore, as sample sizes increase, power also increases. Second, if the change, difference, or pattern observed in the sample is larger or clearer, it is easier to detect and is more likely to represent a difference that would occur in the population than if it is smaller. Thus, as the size of changes, differences, or patterns observed increases, power also increases. Finally, the lower the error is, the closer the observations are expected to be to the parameters of a population. Thus, as the size of error decreases, power increases. Considering these three things together, we can summarize the components that increase power as follows:

1. The larger the sample size, the more closely the sample statistics are expected to represent the population.
2. The larger or clearer the change, difference, or pattern observed in the sample, the more likely it is that it reflects a change, difference, or pattern in the population.
3. The less error there is in the sample statistics used to assess changes, differences, or patterns, the more likely it is that they reflect changes, differences, or patterns in the population.

Obtained Values

These components that impact statistical power are interconnected. Means are estimates for which variability must be considered. Some measures of variability and error (such as standard errors) include sample sizes in their calculations. Further, the greater the sample size, the closer a sample is to being equivalent to its population size. Thus, the formulas used in inferential statistics include variations of some or all of the three components of power to yield one of several forms of obtained values. **Obtained values** are results that summarize data by using inferential formulas. Inferential formulas are those used to test hypotheses. These formulas and other analyses that accompany them take into account the components of power. Data are plugged into inferential formulas which yield obtained values. Those obtained values are compared to specific thresholds to assess whether the data supported or failed to support a hypothesis. Thus, obtained values can be thought of as summaries of how much power or evidence there is to support a hypothesis.

Determining Significance

Statistical significance refers to the determination that a hypothesis is likely true in the population because there is sufficient evidence in the sample to support the hypothesis. Another way to say this is that a statistically significant result occurs when the hypothesized result was observed in the sample with enough power to conclude that the observed result was unlikely to be simply due to random chance. Essentially, when an obtained value is high enough for a given situation, it represents sufficient evidence to declare a hypothesis is significantly supported.

Significance is not absolute. Instead, it is a determination that a hypothesis is *likely* true but not that it is proven to be true. Recall that sampling error is assumed whenever a sample is drawn and used to represent a population. Recall also that there is no guarantee that the sample will represent the population well. Thus, there is always some chance that a hypothesis is not true in the population but that it will appear to be true in the data from a sample. The stronger the evidence is in favor of the hypothesis within the sample, the more likely it is that the hypothesis is true of the population. To say it another way, the stronger the results are, the less likely it is that they would have occurred simply due to random chance rather than because they are true. Therefore, when a result matches a hypothesis and is significant, statisticians conclude that a hypothesis is likely true and, thus, is supported by the evidence. Note that statistical significance is not necessarily indicative that a result is meaningful or useful. Instead, statistical significance simply indicates that the hypothesis is likely true based on the evidence.

Critical Values

Obtained values are compared to critical values to determine whether a hypothesis has enough evidence to be declared significant and, thus, supported. **Critical values** represent thresholds of the minimum amount of evidence that is needed to determine statistical significance and conclude that a hypothesis is supported. Thus, when the obtained value (which represents the amount of evidence) exceeds the critical value (which represents the minimum amount of evidence needed to support a hypothesis), the conclusion is that the hypothesis is supported. Conversely, when the obtained value does not exceed the critical value, the conclusion is that there is insufficient evidence to support the hypothesis. Another way to say this is that the null hypothesis is rejected when the obtained value exceeds the critical value and is retained or accepted when the obtained value does not exceed the critical value.

Obtained and critical values depend on several things which can include whether or not a hypothesis is directional, which inferential formula was used, and the relevant components of power for the hypothesis and corresponding formula used. We will review the specific differences and ways both obtained values and their critical values are found in subsequent chapters. For now, it is only necessary to know that each time a hypothesis is tested, an obtained value must be found, a critical value must be found, and the two must be compared. These are important steps in the larger processes of hypothesis testing.

Steps in Hypothesis Testing

In order to test a hypothesis, these steps should be followed, in the recommended order:

1. State the hypothesis.

This is a necessary first step. Before a study can be designed, a researcher needs to specify exactly what the hypothesis is what they intend to test. Then the process for collecting data (which is the research method) can be developed and carried out, accordingly.

It is worth noting that it is possible to develop a hypothesis after data have been collected but this is not ideal as it introduces important limitations to the research process. Though these limitations are beyond the scope of this book, they are an important topic which is generally covered in a Research Methods course. The focus of this book is best practices for statistical analysis; in keeping, we will always presume a hypothesis was developed before data were collected to test it. Thus, the first step for our analyses and reporting our results will always include stating the hypothesis.

2. Choose the inferential test (formula) that best fits the hypothesis.

There are a variety of formulas, each of which best fits only certain kinds of data and, thus, each only fits certain hypotheses. For example, one test is used to compare the means of the same group at posttest to itself at pretest, a different one is used to compare the mean of one group to the mean of a different group, another is used to compare the means of three or more distinct groups, and still others are used to assess patterns between two or more quantitative variables. The test selected should be the one that is best suited to the hypothesis under investigation. Note: A brief summary of the different kinds of inferential tests included in this book appears towards the end of this chapter.

3. Determine the critical value.

The critical value refers to the number you must surpass in order to conclude that your results are unlikely to be due to chance and, thus, likely reflects a truth about the population. The critical value is a concept we will discuss in more detail in subsequent chapters. For now, know that we weigh the implications of an inaccurate conclusion (e.g. what are the risks of concluding our medication worked when it actually did not) and then set the statistical risk we are willing to take that we might be wrong (which is used to determine the critical value). In the behavioral sciences, we very often decide that we are willing to accept less than a 5% chance that we will conclude a hypothesis is true when it is not; this means we want less than a 5% chance that our result is simply a false positive. Thus, critical values are usually computed to represent the amount (or strength) of evidence that is needed to be at least 95% confident that the hypothesis is true.

4. Calculate the test statistic.

This is the step of the scientific method (and, thus, also in the process of hypothesis testing) in which data are analyzed. In this step, the statistician uses the inferential test that was chosen in step 2 to analyze the data and yield a result. The result is represented by the obtained value (which is also known as a test statistic or result). This is the most math-intensive step of testing a hypothesis.

5. Apply a decision rule and determine whether the result is significant.

In this step, we assess whether our result (i.e. our obtained value or test statistic) exceeds the critical value. When it does, we can conclude that there is a strong probability that the hypothesis is true in the population based on the evidence observed in the sample. In so doing, the result is concluded to be significant. Conversely, when the test statistics does not exceed the critical value, we conclude that the evidence is not strong enough to conclude that the hypothesis is likely true in the population and, thus, that the hypothesis is not supported. In so doing, the result is concluded to be non-significant.

> Note
>
> When a result is close to exceeding the critical value but does not, it may be prudent for researchers to retest the hypothesis or similar hypotheses with new samples in the future. A result which is close to, but does not surpass, the critical value may be referred to as “trending”; however, trending results should not be referred to as significant.

When it is determined that the result is significant, proceed through each of the remaining steps. When it is determined that the result is not significant, skip to step 7 to complete the process of hypothesis testing.

6. Calculate the effect size and other relevant secondary analyses.

An effect size can be reported alongside a significant result. Essentially, an **effect size** is an estimate of the magnitude of an effect, change, or pattern observed in the sample data. Effect size can help statisticians and audiences deduce practical significance. **Practical significance** refers to whether there is a large enough magnitude of effect to be meaningful or useful. This is important because it is possible to have a result that is statistically significant without being practically significant. Thus, it is often recommended that practical significance be reported as a secondary analysis when a result is statistically significant.

Some tests have additional secondary analyses which are necessary to adequately test a hypothesis. In each chapter for which these are recommended, they will be included in the section for step 6 of hypothesis testing.

7. Report the results in American Psychological Associate (APA) format.

Results for inferential tests are often best summarized using a paragraph that states the following:

a. The hypothesis and specific inferential test used,
b. The main results of the test and whether they were significant,
c. Any additional results that clarify or add details about the results, and
d. Whether the results support or refute the hypothesis.

It is recommended that effect sizes be reported with the additional results, when possible and/or common in the field into which the researcher is disseminating results. Dissemination refers to the formal sharing of results which is often done through publishing peer-reviewed, empirical articles in research or academic journals, giving conference presentations, and/ors reporting results in books that focus on summarizing several empirical studies. APA format specifies the level of rounding and types of symbols which should be used when reporting results for each of the various descriptive and inferential tests.

We will employ these steps when we learn how to select and properly use inferential statistics to test hypotheses in the subsequent chapters of this book. In each of those chapters, the details of the formulas, the calculations they require, and the symbols and rounding rules will be covered in detail. It will likely be helpful to refer back to this section with each of those chapters to remind yourself of the order and purpose of each of these steps to testing a hypothesis and reporting the results.

Reading Review 6.3

1. In which step of hypothesis testing are data analyzed?
2. What does statistical significance mean?
3. Which two values are compared to determine whether a result is statistically significant?
4. What is used to estimate practical significance?

6.5: Errors and Statistical Significance

It is important to understand what statistical significance does and does not tell a statistician and how it is determined. Therefore, we will review some important concepts connected to statistical significance in this chapter before learning about the kinds of statistical tests which can have significant results in the subsequent chapters.

There are a few different ways that statistical significance can be understood, all of which are connected to and complement each other. So far we have defined significance in two ways: 1. It refers to the determination that a hypothesis is likely true in the population because there is sufficient evidence in the sample to support the hypothesis and 2. The hypothesized result was observed in the sample with enough power to conclude that it was unlikely to be due to random chance. These are two ways of stating the same overarching concept.

Significance can also be described as the determination that the risk of a Type I Error is sufficiently low. Type I Error is one of two types of error in conclusions that are possible anytime a hypothesis is tested. These two types of conclusion error are tied to something known as an alpha level. Therefore, we must review their connection to get a complete understanding of what it means when we determine that a result is statistically significant.

No Conclusion Error

When a hypothesis is tested, a determination is made as to whether the hypothesis has been supported or not supported. When a result is in favor of a hypothesis and is significant, it is concluded that the hypothesis is likely true. When a result is not significantly in favor of a hypothesis, it is concluded that the hypothesis is too likely to be untrue to be supported. Another way to say this is that when a result is not significantly in favor of a hypothesis, the null hypothesis is retained. When the results from samples and their corresponding conclusions match what is true of the population, there is no conclusion error and the process of hypothesis testing is functioning as hoped.

The Ideal Situation

Ideally, population truths are reflected in sample data allowing consistent, accurate conclusion to be drawn about the population. This can happen one of two ways:

1. The hypothesis is true in the population and is reflected in the sample with sufficient evidence to support the hypothesis and reject the null or,
2. The hypothesis is false in the population and this is reflected in the sample such that the hypothesis is not supported and the null is retained.

In each of these versions of the ideal situation, there is no conclusion error.

Unfortunately, a statistician cannot know if their data from any one sample accurately reflects truths in the population. This is because what is true in a population is unknown. If we could test and know what was true of the population, there would be no need for statistics, but (for better or worse) statistics are necessitated by the fact that we simply cannot always know what is true of populations most of the time. This leaves the possibility of two kinds of conclusion error: Type I Error and Type II Error.

Type I Error

Type I Error is an erroneous conclusion wherein the null hypothesis is rejected when, in fact, it is true of the population. These types of errors are the same as “false positives.” Another way to say this is that a Type I Error occurs when the evidence suggests the hypothesis is true when it is not. This can happened because, though sample data are expected to approximate what is true of populations, samples vary in their sampling error (i.e. how accurately they reflect the population). Thus, false positives can occur. Critical values are set at levels that minimize this risk by requiring that the preponderance of the evidence be in support of the hypothesis before it is concluded that the hypothesis is likely true.

Type II Error

Type II Error is an erroneous conclusion wherein the null hypothesis is retained when, in fact, it is not true of the population. These types of errors are the same as “false negatives.” Another way of saying this is that a Type II error occurs when the evidence suggested the hypothesis was not true when it actually was true. This, like Type I Errors, is possible because of the existence of sampling error. Type II Errors are usually considered preferable to Type I Errors. Therefore, the critical values are set to minimize Type I Error. However, setting critical values to minimize Type I Error causes the risk of a Type II Error to be higher.

Reducing the Risk of Conclusion Errors

The conclusion errors are important because research findings are used to understand the world and to guide actions; it is, therefore, important to consider and reduce the risk of conclusion errors. Statisticians cannot guarantee that neither a Type I Error nor a Type II Error will occur. However, there are a few things that can be done to reduce the risk of one of these two forms of conclusion error. First, researchers should ensure they are using the best measures and methods available to collect data about each variable. Doing so can increase the accuracy of the data being used to test the hypothesis. Second, replication can and should be used. **Replication** is the act of repeating a study under the same or similar conditions and methods as were used in a prior study to see if the same results are obtained. If you have ever searched a research database for articles, you may have noticed that there are often many articles on the same topic, some of which used the same procedures to test the same hypotheses. This is done because each study has the risks of Type I and Type II Error so any one study is often considered insufficient support on its own. Therefore, studies can be repeated to see what the pattern of results is. If the same result keep occurring in the vast majority of studies, it increases the confidence of the scientific community that the conclusions in those studies are correct. A third way to reduce the risk of a conclusion error is to use a stringent threshold for determining significance. The threshold is set using something called an *alpha* level.

Meta-Analyses

Meta-analyses are a unique and very useful form of research. These are used to analyze and summarize the results of many replication studies together. Doing this is favorable over relying on any single study to determine whether a hypothesis is likely true. Thus, when a meta-analysis exists relevant to a topic of interest, it can be a great resource. This is a good thing to keep in mind whenever you are researching a topic for a class paper or even developing your own hypotheses to test. When reviewing empirical research, always check whether a relevant meta-analysis exists.

Alpha Levels

The acceptable probability of a Type I Error is referred to as an alpha level which is represented with the symbol α. The probability of a Type II Error is referred to as a beta level which is represented with the symbol β. Alpha levels are used to set thresholds for significance because the main conclusion error that a statistician is trying to avoid is a Type I Error. Recall that the null hypothesis is presumed to be true by default until sufficient evidence supports rejecting it in favor of the alternative hypothesis. This process prioritizes reducing the risk of a false positive (Type I Error) over reducing the risk of a false negative (Type II Error). In fact, the naming of the errors makes clear which error is the primary one to avoid: Type I.

Statisticians must identify an acceptable alpha level as part of step 3 of hypothesis testing. When a statistician sets an alpha level, they are identifying the risk they are willing to take of a Type I Error if they reject the null and conclude that the alternative hypothesis is supported. Unfortunately, the alpha level cannot be set to 0, which is part of why there are no guarantees in statistics. Setting alpha to 0 would essentially mean that there was no chance of a Type I Error and that a significant result would mean the researcher was 100% sure of this. Though the alpha level cannot be set to 0, it can be set pretty low. If an alpha level is set at .05 (meaning 5%), it would mean that the statistician was taking up to a 5% risk of a Type I Error. This is generally considered an appropriate level of risk. However, if the alpha level was set to .50 (meaning 50%), for example, it would mean that the statistician was taking a 50% risk of a Type I Error; this would mean it could be just as likely that the hypothesis was wrong as that it was right. At that point the researcher might as well save themselves the trouble of collecting and analyzing data and flip a coin to decide if a hypothesis is supported. Thus, a useful and acceptable alpha level is one that is set fairly low.

Most areas of the behavioral sciences consider a 5% risk of a Type I Error to be the maximum acceptable level. Alpha levels are typically written in decimal form so this risk can be summarized as follows: α = .05. This means that a result will be considered non-significant when it is equal to or greater than .05 (i.e. 5%). When this alpha level is used, a researcher must be more than 95% sure that they are not making a Type I Error, based on the power of the evidence, to conclude a hypothesis is supported. This would mean that there is less than a 5% risk of a Type I Error. Sometimes this is worded by saying that the researcher is at least 95% confident they are not making a Type I Error. Another commonly used alpha level is .01. This translates to being at least 99% confident that there is no Type I Error (and a less than 1% risk of a Type I Error).

Though an alpha level cannot be set at 0, it is reasonable to wonder why it isn't set as close to 0 as possible by using alpha levels such as .01 all the time or even .000001 so a statistician could be 99% confident or even 99.9999% confident, respectively. Reducing the alpha level can be advisable in situations where making a Type I Error could have especially problematic consequences. However, the lower a Type I Error is, the higher the risk of a Type II Error becomes. Therefore, decreasing the alpha rate doesn't necessary reduce the overall risk or an erroneous conclusion; instead, it may just be shifting it between the two risks.

Bonferroni correction

An additional consideration for setting alpha levels should be considered when multiple hypotheses are being tested together. Specifically, each time a hypothesis is tested with an alpha level of .05, there is a 5% risk of a Type I Error. That would mean that if two hypotheses were tested in a study, that there would be a 10% chance that at least one of them had a Type I Error, without being able to identify which, if either, had the error. When there are three hypotheses, the risk increases to 15% and so on. Therefore, when multiple tests are used, statisticians are advised to consider reducing the alpha level for each test to keep the overall risk of a Type I Error for the tests together below 5%.

The commonly used correction for this is known as a Bonferroni correction. The application of this method for addressing increased risks of Type I Error is attributed to a biostatistician named Olive Jean Dunn (Dunn, 1961; though it is named for the mathematician Carlo Emilio Bonferroni who focused on foundational issues of probability). To adjust risk of a Type I Error using a **Bonferroni correction**, the alpha level (α) is divided by the number of tests used (m) to get the reduced alpha level to be used for each test. For example, if the researcher wants a total risk of Type I Error to be less than 5% (α = .05) when two hypotheses are tested (m = 2), the alpha level that would be used for each test is .025 or 2.5% (because $\frac{\alpha}{m} = \frac{.05}{2} = .025$).

Reading Review 6.4

1. In statistics, a false positive is referred to as which type of error?
2. In statistics, a false negative is referred to as which type of error?
3. Which alpha level is most commonly used?
4. When and why would a Bonferroni correction be recommended?

6.6: Choosing Statistical Tests

Inferential statistics are used to test hypotheses and build upon and beyond descriptive statistics. Descriptive statistics can be used to summarize things such as counts, means, and standard deviations, yet they are unable on their own to tell you whether differences in any of those are likely true beyond the samples whose data were being summarized. Inferential techniques take descriptive statistics such as these into further calculations to estimate whether differences or patterns in them are strong enough to indicate a likely difference or pattern in the broader population. Therefore, inferential statistics are not fully distinct from descriptive statistics; instead, they subsume various descriptive statistics into more complex calculations to establish the probabilities that they are indicative of population truths.

Breaking Down a Hypothesis for Testing

There are many different inferential techniques that can be used to test hypotheses. Each inferential test fits specific kinds of hypotheses and their corresponding data. Inferential testing starts with the statement of a hypothesis. Thus, a hypothesis must be stated before an appropriate inferential test can be chosen to test it. Once stated, a hypothesis can be broken down to identify:

1. How many variables are included
2. How many of those variables are quantitative (such as scores) or qualitative (such as group names)
3. Which of those variables, if any, are considered independent variables and which, if any, are considered dependent variables
4. The proposed nature of those variables in regards to one another

Let's take a look at a hypothesis and walk through each of these. Suppose that it is hypothesized that:

> Cholesterol levels will be lower for those who ate oatmeal daily for six weeks compared to those who did not.

Number of Variables

First, identify how many variables are in the hypothesis. Remember, a variable is anything that is measured and is not always the same. There are two proposed variables in this hypothesis:

1. Cholesterol Level and
2. Oatmeal Consumption.

Types of Variables

Second, identify whether each of the variables is quantitative or qualitative. Cholesterol levels are quantifications of how much cholesterol is circulating in the bloodstream. Thus, cholesterol level is a quantitative variable. Oatmeal consumption is qualitative in this example. This is because members of the sample are being grouped based on whether they did or did not consume the oatmeal for the specified timeframe. Comparisons of one group to the other are going to be made. However, you may ask yourself whether oatmeal consumption can be quantitative, and rightly so because it can be in other hypotheses. It is possible to measure the *amount* of oatmeal each person consumed rather than to identify simply whether or not they had consumed it for a specified time frame. However, what matters when identifying variables to choose an inferential tests is not how a variable *could* be measured but, instead, how it is being identified for measurement in the current hypothesis. In the aforementioned research hypothesis, it is clear that oatmeal consumption is being measured qualitatively. Therefore, for this hypothesis, there is one quantitative variable (cholesterol level) that is continuous in nature and one qualitative variable (oatmeal consumption).

Third, differentiate between the independent variable(s) and the dependent variable(s). Recall that an independent variable is the theorized or hypothesized cause and the dependent variable is the theorized or hypothesized affected in a cause-effect relationship (for a review, see Chapter 1). Not all studies have variables which are true independent variables and dependent variables. Instead, some hypotheses are about whether a pattern or difference exists, without positioning variables as causes and effects in relation to one another. However, it is necessary to distinguish between them when applicable. In addition, sometimes a theory on which a hypothesis was generated specifies cause-effect even when the hypothesis does not do so overtly or the data collected cannot be used to deduce cause-effect. We will return to these caveats and expand on them in later chapters where they are most applicable. For now, let's continue breaking down the components of our example hypothesis.

Though causal language is not used in the hypothesis, it may be implied by the structure of the hypothesis. Oatmeal consumption is occurring first, over the course of six weeks, before cholesterol levels are measured in the two groups. Thus, oatmeal consumption is being positioned as though it is an independent variable and cholesterol is being positioned as though it is a dependent variable.

Thus, if the method of data collection is in keeping with this, oatmeal consumption is a qualitative, independent variable and cholesterol is a quantitative, dependent variable.

Proposed Relationship between Variables

Finally, we must identify the nature of the connection, pattern, or distinction between the variables which is being proposed in the hypothesis. Though there are many forms of relationships that can be proposed and tested, some of which can be quite complex, the focus of this section will be limited to the forms of tests covered in this book.

Common relationships that can be proposed and tested are:

Group Differences

1. That the counts of groups are different
2. That the means of groups are different

Patterned Connections

3. That variables trend, or form a pattern, together
4. That established trends/patterns can be used to predict connections in new data

Let's first decide which category best fits the proposed relationship in the aforementioned hypothesis. We have a grouping independent variable so we turn to the Group Differences options. Now we need to decide whether our other variable quantifies how many cases have cholesterol (option 1) or whether cholesterol is being measured in each case allowing a mean cholesterol level to be computed for each group (option 2). Amount of cholesterol is being measured for each case. As we saw in Chapter 3, quantitative variables are often summarized using means. Thus, mean cholesterol scores would be computed for each group. Together this means that option 2 (That group means are different) is the best categorization for the proposed hypothesis.

It is often useful to create a summary as you progress through any processes in research and statistics to keep track of all the necessary information. In keeping, here is a summary of the hypothesis and how we have broken it down:

Hypothesis: Cholesterol levels will be lower for those who ate oatmeal daily for six weeks compared to those who do not.

Variables	Variable Type	Variable Position (if any)
1. Cholesterol Level	Quantitative (continuous)	Potential DV
2. Oatmeal Consumption	Qualitative	Potential IV

Proposed Relationship: The means of groups are different.

> **Note**
>
> Hypotheses are often presented in future tense before data are collected and in past tense after data have been collected and analyzed.

Summary of Inferential Tests

Once a hypothesis has been broken down, we can review the options for inferential testing and choose the one that best fits the hypothesis.

Group Differences

Each of the group differences tests requires that the data set and hypothesis include at least one grouping (qualitative) variable.

1. That the counts of groups are different
    a. To compare the counts of cases or objects in different groups either to each other or to other specified counts: Chi-Squared Goodness of Fit Test
    b. To test whether counts in one group appear to be dependent upon counts in another group: Chi-Squared Test of Independence
2. That the means of groups are different
    a. To compare a sample mean to the known or hypothesized population mean: One-Sample *t*-Test

b. To compare the means of two different groups to each other: Independent Samples *t*-Test
c. To compare the means of three or more different groups to one another: One Way ANOVA (also known as Simple ANOVA or Independent Groups ANOVA)
d. To compare one group to itself at two time points: Dependent Samples *t*-Test (also known as a Paired Samples of Repeated Measures *t*-test
e. To compare one group to itself at three or more time points: Repeated Measures ANOVA

Patterned Connections

Each of the patterned connections tests requires that the data set and hypothesis include two quantitative variables.

3. That variables trend, or form a pattern, together
   a. To test the relationship between two quantitative variables: Bivariate Correlation
4. That established trends/patterns can be used to predict connections in new data
   a. To test whether one quantitative variable is useful for predicting another quantitative variable: Simple Linear Regression

Though there are many variations beyond these tests, these are the foundational inferential statistics used by many social and behavioral researchers and, thus, will be the ones covered in this book. Each of these tests is covered in its own chapter with the exception of the two Chi-Squared test which are covered together in the final chapter of this book.

Reading Review 6.5

Complete each table using the hypotheses provided above it.

1. **Hypothesis:** People will have greater mean hours of sleep after they stop consuming caffeine compared to before.

Variable Name	Variable Type	Variable Position (if any)
Which inferential test is the best fit for this hypothesis?		

2. **Hypothesis:** Level of happiness will be related to income.

Variable Name	Variable Type	Variable Position (if any)
Which inferential test is the best fit for this hypothesis?		

3. **Hypothesis:** Students who are given study guides will have higher mean test scores compared to those who are not given study guides.

Variable Name	Variable Type	Variable Position (if any)
Which inferential test is the best fit for this hypothesis?		

6.7: References

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American statistical association*, *56*(293), 52-64.

CHAPTER OVERVIEW

7: One Sample t-Tests

7.1: t-Statistics and the One Sample t-tests

The first group of inferential tests covered in this book are the *t*-tests. There are three kinds of *t*-tests: the one sample *t*-test, the dependent samples *t*-test, and the independent samples *t*-test. Each of these is best suited for comparing the means of two groups of data under different conditions. What the *t*-tests have in common is that each compares the means of two groups of scores for the same variable. They differ in what makes up the two groups of scores. The one sample *t*-test is used to compare the mean of a sample to a known or hypothesized population mean. The dependent samples *t*-test is used to compare the mean of a single sample at one time to itself at another time. The independent samples *t*-test is used to compare the mean of one sample to the mean of a different sample. Each of these uses a different formula but all require the use of sample data. Each of these is the focus of one chapter, starting with the one sample *t*-test for this chapter.

One Sample t-Test

The one sample *t*-test is used to test whether the mean from a sample is significantly different from another hypothesized mean such as a known population mean. When the population parameter is known, we can assess how similar our sample statistic is to the population parameter. This makes the one sample *t*-test a bit usual because one of its common applications can only be used when the population parameter is known yet statisticians typically use sample estimates *because* the population parameter has not been or cannot be directly measured. However, this version of the one sample *t*-test does have utility; it can be used to estimate how representative a sample is for its population. For example, say you wanted to conduct research to see how a sample of students felt about their school but that you had concerns that they may not represent the population of students well if they had a significantly different grade point average (GPA) than the student body at the school, broadly. A one sample *t*-test could be conducted first to assess whether the sample's GPA was similar or significantly different from the overall student GPA at the school. If the sample's mean GPA is not significantly different from the known mean GPA of the full population of students at the school, the researcher may comfortably proceed to collecting and analyzing data about other variables in the sample to understand the feelings of the corresponding population of students.

A second use of a one sample *t*-test is to see if a sample mean is similar or significantly different from a hypothesized value for the mean, whether or not that value is known to be true in a population. In this version, a hypothesized or hoped for value is used in place of a known population mean. For example, suppose that a nail polish manufacturer has a machine that is supposed to distribute 15 ml of product into each bottle and that the owner wants to check whether the machine is depositing amounts appropriately. To do so, the owner randomly samples 100 bottles filled by the machine and measures the amount of polish in each bottle. If the machine is working properly, the mean volume for the sample should *not* be significantly different from 15 ml. This scenario can also be tested using a one sample *t*-test.

Data and Hypotheses that Fit the One Sample t-Test

Data

Each statistical test has some assumptions which must be met in order for the formula to function properly. In keeping, there are a few assumptions about the data which must be met before a one sample *t* test is used. First, the one sample *t*-test is univariate. This means that only one variable is being tested or analyzed at a time using data from a sample. Second, the variable should be measured on the interval or ratio scale. If these conditions are not met, the one sample *t*-test should not be used. Third, and warranting further explanation, the data must be independent. Let's take a moment to get a clear understanding of what that means before moving on to a discussion of hypotheses.

Independent Data

The one sample *t*-test is a univariate analysis that is suited for independent data. The word *independent* is used in several ways in statistics, each time distinguishing a particular attribute of a variable or data. When data for a variable are independent, it means that the scores for each case are not dependent upon, directly connected to, or influential of one another. For example, if data are collected about hours of sleep from a random sample of people, the data are likely independent. The number of hours slept by person 1 do not impact how many hours person 2 or person 3 slept. In this example, the scores for the variable Hours of Sleep are independent of each other. In contrast, the data for Hours of Sleep would *not* be independent for two new parents who have decided to sleep in shifts so that one is awake while the other sleeps, and visa-versa. In this example the hours parent 1 sleeps are dependent upon the hours parent 2 sleeps. This is an important consideration when working with univariate data and statistical tests. Hours of sleep could be analyzed using a one sample *t*-test for the random sample of independent scores for sleep hours but not for data from the couples with dependent sleep hours. When data for a variable are independent, the one sample *t*-test can be used.

Hypotheses

Hypotheses for the one sample *t*-test can be non-directional or directional. For the one sample *t* test, the non-directional research hypothesis is that the sample mean will be different from the population mean (or an otherwise hypothesized mean). The corresponding null hypothesis is that the sample mean will not be different from the population mean (or an otherwise hypothesized mean). Because this research hypothesis is non-directional, it requires a two-tailed test. The non-directional research and corresponding null hypotheses can be summarized as follows:

Non-Directional Hypothesis for a One Sample t-Test

Research hypothesis	The sample mean is not equal to the population mean (or hypothesized mean)	$H_A: \bar{X} \neq \mu$
Null hypothesis	The sample mean is equal to the population mean (or hypothesized mean)	$H_0: \bar{X} = \mu$

There are two directional hypotheses possible for the one sample *t*-test. One possible directional research hypothesis is that the sample mean will be *higher than* the population mean (or an otherwise hypothesized mean). The corresponding null hypothesis is that the sample mean will *not* be higher than the population mean (or an otherwise hypothesized mean). This could mean that the sample mean is less than or that it is equal to the population mean. Because this research hypothesis is directional, it requires a one-tailed test. This version of the research and corresponding null hypotheses can be summarized as follows:

Directional Hypothesis for a One Sample t-Test stating Sample Mean will be Higher

Research hypothesis	The sample mean is higher than the population mean (or hypothesized mean)	$H_A: \bar{X} > \mu$
Null hypothesis	The sample mean is not higher than (i.e. is lower than or equal to) the population mean (or hypothesized mean)	$H_0: \bar{X} \leq \mu$

For the one sample *t*-test, the other possible directional research hypothesis is that the sample mean will be *lower than* the population mean (or an otherwise hypothesized mean). The corresponding null hypothesis is that the sample mean will *not* be lower than the population mean (or an otherwise hypothesized mean). This could mean that the sample mean is greater than or that it is equal to the population mean. Because this research hypothesis is directional, it requires a one-tailed test. This version of the research and corresponding null hypotheses can be summarized as follows:

Directional Hypothesis for a One Sample t-Test stating Sample Mean will be Lower

Research hypothesis	The sample mean is lower than the population mean (or hypothesized mean)	$H_A: \bar{X} < \mu$
Null hypothesis	The sample mean is not lower than (i.e. is higher than or equal to) the population mean (or hypothesized mean)	$H_0: \bar{X} \geq \mu$

These three version of the hypothesis are the broad form and would be refined to include the specific name of the variable that is under investigation when working with specific hypotheses. For example, in the nail polish bottling example, a non-directional hypothesis would be used to test whether the sample bottles were different than the intended 15 ml each if there was no reason to believe that the machine was over-filling or under-filling, specifically. This version of the research and corresponding null hypotheses can be summarized as follows:

Specific, Non-Directional Hypothesis for a One Sample t-Test

Research hypothesis	The sample mean volume of nail polish is not equal to 15 ml.	$H_A: \bar{X} \neq \mu$
Null hypothesis	The sample mean volume of nail polish is equal to 15 ml.	$H_0: \bar{X} = \mu$

In this example, the owner and/or manufacturer is testing the research hypothesis but is likely hoping that, in fact, the null hypothesis is retained. This is because they likely hope that the machine is working as intended (by depositing 15 ml per bottle). You may reasonably wonder why, then, the research hypothesis is not that the sample mean *is* equal to 15 ml instead of the null hypothesis stating this. This is because the default when testing is that there is nothing to find: there is no difference, no change, and/or no pattern. That is, the null statement that there is nothing to find (e.g. there is no difference) is assumed and is only rejected when data indicate otherwise. Formulas and the steps to hypothesis testing require that there is strong enough evidence against this to reject the null and, thus, to be able to conclude that a difference, change, or pattern likely exists. If we believed there was truly no difference to find, no data would need to be collected or tested because the null would already be accepted. However, when we are unsure or believe there is a difference, data need to be collected and tested. The default null assumption that there is nothing to find (no difference, change, or pattern) is foundational to how hypothesis testing is carried out and guarding against the risk of a Type I Error (see Chapter 6 for a review of Type I Error).

7.2: The Logic of a t-Distribution

Testing hypotheses using a *t*-distribution is similar to using a z-score distribution. In each of these distributions, samples are drawn from the populations they are meant to represent. When samples are drawn and used repeatedly to create summary statistics, those statistics together will tend to approximate the normal curve when either the distribution of population values are normal or if the sample sizes are sufficiently large. The larger the sample size, the closer the *t*-distribution will tend to be to a normal distribution. It is also important to note that *t*-statistics, like many other statistics, are intended for data which are drawn from the population through the process of random sampling. When data are not randomly sampled, the estimates yielded may be less reliable. Therefore, caution should be used when working with data which are not from random samples, despite the fact that these kinds of samples are quite common in the behavioral and social sciences.

Reading Review 7.1

1. What is the general hypothesis that can be tested using a two-tailed, one sample *t*-test? Provide both sentence and symbol formats.
2. How many variables can be tested each time a one sample *t*-test is used?
3. In what ways is a *t*-distribution similar to a z-distribution?

7.3: The One Sample t-Test Formula

The formula is set up as a division problem with four symbols which represent and, thus, must be replaced with specific values. The numerator focuses on the difference between the sample statistic and the (known or hypothesized) population parameter for a variable. This is the core of the formula because the hypotheses tested using a one sample *t*-test are specifically asking whether these two values differ. Therefore, you can think of the numerator as the main focus of the formula and the denominator as taking into account other necessary information and adjustments. This will be true for the main format used for all three versions of the *t*-test in this book. The denominator of the one sample *t*-test formula is used to take into account the error in the sample (via the standard deviation) and the sample size. The one sample *t*-test formula is as follows:

$$t=\frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

Notice that the formula, though small, includes all three components that impact statistical power (see Chapter 6 for a review of the components of power):

1. The size of the change, difference, or pattern observed in the sample
2. The sample size
3. The size of the error in the sample statistic(s)

The first component, size of difference, is being incorporated in the numerator. The second and third components of sample size and error are being incorporated in the denominator.

Let's take a moment to further examine the denominator. Recall that we expect some error when we draw a sample from a population and that we call this sampling error. We often estimate this error with the formula for standard error (SE). SE is calculated by dividing the sample standard deviation by the square root of the sample size (i.e. $SE=s/\sqrt{n}$). Thus, the denominator for a one sample *t*-test is actually the standard error formula (see Chapter 6 for a review of standard error). This means that we have the SE formula *inside* of our one sample *t*-test formula. Smaller formulas appearing inside larger formulas will happen a lot as we move though the various inferential statistics formulas. This is because statistics builds on some underlying estimates and concepts that get subsumed into more complex formulas to address different needs. If you keep this in mind, it can make future formulas easier to learn and understand, especially when the same formula may be written a few different ways. For example, the one sample *t*-test formula can be written as follows:

$$t=\frac{\bar{x}-\mu}{SE}$$

This version of the formula requires the same information and steps as the one shown previously; the only difference is that the steps used to calculate the standard error in the denominator have been replaced with the symbol for *SE*. Therefore, these two versions of the formula have the exact same steps and will yield the same result. You can use either version. However, this abbreviated one requires that you either remember or look up how to calculate SE. In contrast, the previous version shows you how to calculate SE. Therefore, I recommend using the previous one showing the SE formula rather than the latter one showing the SE symbol, especially if you are newer to statistics and/or to memorizing mathematical formulas and symbols in general.

Study Tip

When adding formulas to your notes:

1. use the version which shows more steps and
2. note for yourself when a group of symbols and operations within a formula are actually another formula altogether.

This will help you connect across topics and see that much of statistics breaks down to the same elements repeating in various ways to create new formulas.

One Sample *t*-Test Formula

$t=\frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n}}\right)}$	⇐The Standard Error (SE) Formula appears as the denominator of the one sample t-test formula

$t=\frac{\bar{x}-\mu}{s/\sqrt{n}}$	The formula may also be written like this where the parenthesis for the denominator is implied. The formula and steps are the same as for the version shown above.

Formula Components

In order to solve for *t*, four things must first be known:

μ = the hypothesized or known population mean

$\bar{x}$= the sample mean

s = the sample standard deviation

n = the sample size

Some of these components can be given while others generally need to be calculated from a data set. Typically, the information for the population is given and the information for the sample is calculated from a data set when using this formula. If all four components are given, one can proceed to plugging them into the formula and finding the *t*-value using the order of operations. However, because the sample statistics are typically found using data, all steps are shown below assuming these need to be calculated.

Formula Steps

The steps are shown in order and categorized into two sections: A) preparation and B) solving. I recommend using this categorization to help you organize, learn, and properly use all inferential formulas. Preparation steps refer to any calculations that need to be done before values can be plugged into the formula. For the one sample *t*-test this includes finding the descriptive statistics that make up the four components of the formula (listed above). Once those are known, the steps in section B can be used to yield the obtained value for the formula. The symbol for the obtained value for each *t*-test is *t*. Follow these steps, in order, to find *t*.

Section A: Preparation

1. Identify μ using the known population mean or otherwise hypothesized value.
2. Find n using the sample data for the focal variable.
3. Find $\bar{x}$ using the sample data for the focal variable.
4. Find s using the sample data for the focal variable.

Section B: Solving

1. Write the formula with the values found in section A plugged into their respective locations.
2. Solve the numerator by subtracting the population (or hypothesized) mean from the sample mean.
3. Solve for the denominator by finding the square root of the sample size and then dividing the sample standard deviation by that value.
4. Solve for *t* by dividing the numerator (result of step 2) by the denominator (the result of step 3).

You may have noticed that we are solving the numerator before the denominator despite the fact that the denominator is shown in parentheses. This may seem like it goes against the order of operations commonly referred to as PEMDAS. PEMDAS states the correct order of computations is: Parentheses, Exponents, Multiplication and Division, and finally, Adding and Subtracting. However, the P in PEMDAS is better stated as G for Grouping (making the acronym GEMDAS). This is because we work through any groups of operations first, regardless of whether they are in parentheses, brackets, or other grouping structures. Our formula actually has two groups of steps: those in the numerator and those in the denominator. This is because the division sign in the middle of the formula separates the operations into these two distinct groups. Therefore, when we see a horizontal division line (like we see in the one sample *t*-test formula), it is treated as a grouping structure. You can imagine that the whole numerator is within parentheses just like the denominator. For more review of order of operations, see Appendix B.

Data Set 7.1

Test Scores

Test Scores
85
80
80
75
75
75
70
70
70
70
65
65
65
65
65
60
60
60
60
55
55
55
50
50
45

Example of How to Test a Hypothesis by Computing t

Let us assume that a school has implemented a new teaching strategy, and we want to assess whether the students at that school are scoring significantly differently on a standardized test than students in general in the United States. Let us assume that the mean for the population of students in the U.S. is 58.00. Assume that Data Set 7.1 includes data from the school sample. Let's use this information to test a hypothesis.

Steps in Hypothesis Testing

In order to test a hypothesis, we must follow the steps for hypothesis testing:

1. **State the hypothesis**.

A non-directional hypothesis is the best fit because the goal is to see if those at the school had a *different* mean score from the population without having enough information to hypothesize whether the sample will be specifically higher or lower than the population. A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that the mean test score for the school sample will be different from the population mean of 58.00.* If the format shown in the table below is used instead, it must be made clear that what is being stated is a research hypothesis not a result (hence, you see it labeled to the left of the hypothesis as such).

Non-Directional Hypothesis for a One Sample t-Test

Research hypothesis	The mean test score for the school sample will be different from the population mean of 58.00.	$H_A : \bar{X} \neq \bar{\mu}$
Null hypothesis	The mean test score for the school sample will not be different from the population mean of 58.00.	$H_0 : \bar{X} = \bar{\mu}$

2. **Choose the inferential test (formula) that best fits the hypothesis.**

A sample mean is being compared to a population mean so the appropriate test is a one sample t-test.

3. **Determine the critical value.**

In order to determine the critical value, three things must be identified:

1. the alpha level,
2. whether the hypothesis requires a one-tailed test or a two-tailed test, and
3. the degrees of freedom for the test (df).

An alpha level refers to the risk of a Type I Error that is being taken and is summarized with the symbol α. Alpha levels are often set at .05 unless there is reason to adjust them such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The hypothesis is non-directional so a two-tailed test should be used.

The *df* must also be calculated. Each inferential test has a unique formula for calculating df. See the section titled "Deeper Dive: What are Degrees of Freedom?" later in this chapter to learn what df represents and how it is calculated for the different *t*-tests before returning to this section to complete the steps in hypothesis testing. In short, the formula for df for a one sample *t*-test is as follows: $df = n – 1$. The sample size for Data Set 7.1 is 25. Thus, df = 25 – 1 so the df for this scenario is 24.

These three pieces of information are used to locate the critical value for the test. The full tables of the critical values for *t*-tests are located in Appendix D. Below is an excerpt of the section of the *t*-tables that fits the current hypothesis and data. Under the conditions of an alpha level of .05, a two-tailed test, and 24 degrees of freedom, the critical value is 2.064.

Critical Values Table

	two-tailed test	
alpha level:	α = 0.05	α = 0.01
Degrees of Freedom:	24	24
Critical Values:	2.064	2.797

The critical value represents the absolute value which must be exceeded in order to declare a result significant. Think of this as the threshold of evidence needed to be confident a hypothesis is true and think of the obtained value (which is called *t* in a *t*-test) as the amount of evidence present.

4. **Calculate the test statistic.**

A test statistic can also be referred to as an obtained value. The formula needed to find the test statistics *t* for this scenario is as follows:

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

Section A: Preparation

Start each inferential formula by identifying and solving for the pieces that must go into the formula. For the one sample *t*-test, this preparatory work is as follows:

1. Identify μ using the known population value or the hypothesized value.
   - *This value is given as* μ = 58.00
2. Find n using the sample data for the focal variable.
   - *This value is found using Data Set 7.1 and is summarized as* n = 25
3. Find $\bar{x}$ using the sample data for the focal variable.
   - *This value is found using Data Set 7.1 and is summarized as* $\bar{x}$= 65.00

4. Find s using the sample data for the focal variable.
   - *This value is found using Data Set 7.1 and is summarized as* s = 10.2062

> **Note**
>
> For review of how to calculate a sample mean, see Chapter 3. For review of how to calculate a sample standard deviation, see Chapter 4. The standard deviation is shown rounded to the ten thousandths place.

Now that the pieces needed for the formula have been found, we can move to Section B.

Section B: Solving

Now that the preparatory work is done, the formula can be used to compute the obtained value. For the one sample *t*-test, this work is as follows:

Write the formula with the values found in section A plugged into their respective locations. Writing the formula first in symbol format before filling it in with the values can help you recognize and memorize it. Here is the formula with the symbols:

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

Here is the formula with values filled into their appropriate locations in place of their symbols:

$$t = \frac{65.00 - 58.00}{\left(\frac{10.2062}{\sqrt{25}}\right)}$$

Once the symbols have been replaced by values, it is easier to see the mathematical operations which should be followed using the order of operations.

1. Solve the numerator by subtracting the population (or hypothesized) mean from the sample mean.

$$65.00 - 58.00 = 7.00$$

2. Solve for the denominator by finding the square root of the sample size and then dividing the sample standard deviation by that value.

$$\begin{aligned} &\frac{10.2062}{\sqrt{25}} \\ &= \frac{10.2062}{5} \\ &\approx 2.0412 \end{aligned}$$

3. Solve for *t* by dividing the numerator (result of step 2) by the denominator (the result of step 3).

$$\begin{aligned} t &= \frac{7.00}{2.0412} \\ t &\approx 3.4294 \end{aligned}$$

This result, known as a test statistic or *t*-value, can also be referred to by the general term "obtained value."

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value for *t* exceeds the critical value as follows:

The critical value is 2.064.

The obtained *t*-value is 3.4294

The obtained *t*-value does exceed (i.e. is greater than) the critical value, thus, the result is significant.

> **Note**
>
> Only the size of the values, not whether they are positive or negative, is considered when a hypothesis is non-directional (i.e. when a two-tailed test is being performed). Thus, it is the absolute value of t that is being compared to the critical value. However, when a directional hypothesis is used, both the direction and the size of the t-value must be considered.

6. **Calculate the effect size and/or other relevant secondary analyses.**

When it is determined that the result is significant, effect sizes should be computed. Because the result was determined to be significant in step 4, the effect size is needed before proceeding to step 7 to complete the process.

The effect size that is appropriate for t-tests is known as Cohen's d (Cohen, 1988). The formula for Cohen's d is as follows:

$$d = \frac{\bar{x} - \mu}{s}$$

You may notice how similar the effect size formula is to the one sample t-test formula. Cohen's d, when used for a one sample t-test, calculates how many sample standard deviations the sample mean is from the population mean (or hypothesized mean). Thus, the numerator finds the difference in the two means and the denominator is used to divide that by the sample standard deviation. The calculations for the current data would be as follows:

$$d = \frac{65.00 - 58.00}{10.2062}$$
$$d = \frac{7.00}{10.2062}$$
$$d \approx 0.6859$$

Effect sizes, like most values, are rounded and reported to the hundredths place. Thus, this effect size is reported as d = 0.69. Cohen's d can be interpreted using the following rules of thumb (Cohen, 1988; Navarro, 2014):

Interpreting Cohen's d Effect Sizes

~0.80	Large effect
~0.50	Moderate effect
~0.20	Small effect

The rules of thumb are general guidance and do not dictate precise or required interpretations. Instead, they provide some generally agreed upon approximations to aid in interpretations. As is true of all analyses, it is best to consider their situated (or practical) relevance. However, the rules of thumb are useful in providing an initial guideline for interpreting effect sizes. Following these rules of thumb, the current finding of d = 0.69 would be considered a moderate to large effect.

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

Keep in mind that results are reported in past tense because they report on what has already been found. In addition, the research hypothesis must be stated but the null hypothesis is usually not needed for summary paragraphs because it can be deduced from the research hypothesis. Finally, APA format requires a specific format be used for reporting the results of a test. This includes recommendations for rounding and a specific format for reporting relevant symbols and details for the formula and data used. Throughout this book, decimal numbers will be rounded to the hundredths place when reported in a summary sentence or paragraph. Following this, the results for our hypothesis with Data Set 7.1 can be written as shown in the summary example below.

APA Formatted Summary Example

A one sample t-test was used to test the hypothesis that the mean test score for the school sample would be different from the population mean of 58.00. Consistent with the hypothesis, the mean test score for the sample (M = 65.00; SD = 10.21) was significantly different than the mean for the population, $t(24) = 3.43$, $p < .05$. The Cohen's d effect size of 0.69 was moderate to large.

This succinct summary in APA format provides a lot of detail and uses specific symbols in a particular order. To understand how to read and create a summary like this, let's take a detailed walk-though of each piece, what it means, and why it appears where it does. Each inferential test will use this structure, though as we progress through some chapters, the summary paragraph will grow to accommodate the increasing complexity and/or details of the analyses. Thus, in the same way it is necessary to understand the smaller formulas because each chapter builds from them, it is also necessary to understand the basic parts of a summary paragraph because these, too, will be built upon.

7.4: Reporting Results

Reporting results for inferential statistics are much more complex than for descriptive statistics. Where descriptive statistics are usually reported in one sentence using one to two symbols at a time, inferential statistics include both descriptive statistics as a foundation and the results of inferential tests which require additional values and their related symbols. American Psychological Association (APA) format provides useful guidance to ensure that these are reported uniformly across studies by different researchers and statisticians. This helps to both make it easy for other researchers and statisticians to quickly find the information they are looking for, but also facilitates comparisons of results across studies and reduces the chances of misinterpretation caused by formatting differences.

Structure for APA-Formatted Results

In your APA write up for a one samples *t*-test you should state:

1. Which test was used and the hypothesis which warranted its use.
2. Whether the aforementioned hypothesis was supported or not. To do so properly, three components must be reported:
   a. The mean and standard deviation for the sample
   b. The test results as $t(df)$ = obtained value
   c. The significance as $p < .05$ if significant or $p > .05$ if not significant

I will refer to component 2a as the descriptive statistics and components 2b and 2c together as the evidence string with the determination of significance in the following sections. Let's take a detailed look at both parts of an APA write-up and the components of part 2 of the write-up in detail.

Statement of Test Used and Hypothesis Tested

Your job in creating a results paragraph is to tell your reader the story of those results in a clear and succinct manner. The hypothesis is the catalyst for all the work which was done; thus, the start of the story is the hypothesis and the specific inferential test that went with it. It can be tempting to jump right to the results, but that is like starting a movie at the climax and expecting someone to know what is going on and why it is happening. Therefore, the first sentence of a results section should state these two things together: the hypothesis and statistical test used. Though it is possible and permissible to use two sentences to report this information, there are two drawbacks to doing so: 1. You miss the opportunity to emphasize their connection and 2. You are using more words and punctuation than needed to get the job done well. Therefore, let's report the hypothesis and test used for Data Set 7.1 together in a single sentence. Here is an example of how that can be done succinctly:

> A one sample t-test was used to test the hypothesis that the mean test score for the school sample would be different from the population mean of 58.00.

Evidence String for **t-Tests: Results and Significance**

When reporting the results of an inferential test, more than just the obtained value is needed. We must also provide two additional summaries relevant to the test: the degrees of freedom and the significance using something called a p-value. These details are used together in a prescribed format to create what I will refer to as the evidence string. The evidence string goes at the end of the sentence stating the main results and incudes, in this order: the symbol for the inferential test used, the degrees of freedom for the test inside parentheses, the obtained value, and the significance reported using a p-value.

Anatomy of an Evidence String

Symbol for the test	Degrees of Freedom	Obtained Value	p-Value
t	(24)	= 3.43,	$p < .05$.

Symbol for the test

The start of the evidence string provides the symbol for formula used. When a *t*-test is used, a lowercase letter "t" is written in italics to start the evidence string.

Degrees of Freedom

The degrees of freedom (df) are reported inside parentheses. The df summarize how many pieces of evidence (or data points) are able to either help support of refute a hypothesis. The df is calculated as the amount of data (which is equal to the sample size) minus the number of constraints placed on those data when the one sample *t*-test formula is used. Thus, the df can be thought of as an adjusted version of the sample size. For the one sample *t*-test there is one constraint so the df is equal to $n – 1$. Thus, the degrees of freedom also allow the reader to deduce the sample size used for the test (because df is an adjusted version of the sample size). For a more detailed review of df and *t*-tests, see the section titled "Deeper Dive: What are Degrees of Freedom?" later in this chapter.

Obtained value

The main result is known as the test statistic or obtained value. It is what is yielded by using a formula to analyze the data. The obtained value is the result of step 4 of hypothesis testing. It should be rounded and shown to the hundredths place.

p-values

***p*-values** summarize the chance of a Type I Error when concluding that a research hypothesis is supported based on both the evidence (i.e. the data and corresponding obtained value) and the alpha level chosen. The p-value tells two things: which alpha level was used and whether the result was significant when using that alpha level. Think of the letter *p* standing for "*p*robability of a Type I Error." A Type I Error refers to when a hypothesis is supported based on sample evidence but is actually incorrect (see Chapter 6 for a detailed review of Type I Error). When a result is significant the following symbols are used: p < .05. Let's translate each symbol and value then put it together: p stands for "the probability of a Type I Error"; < translates to "less than"; .05 is the 5% alpha level written in decimal format. Taken together, p < .05 can be read as "The probability of a Type I Error is less than 5%." This is just another way of saying that a result is significant. This set of symbols gets attached to the end of any sentence of a results section which declares that a hypothesis was significantly supported. Conversely, when a result is not significant, the symbols are changed and reported as p > .05; this version translates to "The probability of a Type I Error is greater than 5%."

These symbols are used anytime a test for significance (i.e. an inferential statistic) is performed. The fact that these symbols are always used and that they are reported at the end of sentences makes it very easy for researchers and statisticians to scan through a results section and identify which results were or were not significant and, thus, which hypotheses were or were not supported. This is especially helpful when a study includes many hypotheses and tests, which can sometimes span a full page or more. Luckily for us, this is an introductory book, so we get to focus on just one hypothesis and its results at a time. This allows us to practice the use of symbols and where they would be located in a results sentence or paragraph one at a time until they are familiar and comfortable. After familiarity and comfort comes utility and ease and, for the focused statistician, potentially a rewarding future career in research with a nice paycheck.

Reporting p-Values

When a result is significant, the following symbols are used: p < .05.

When a result is not significant, the following symbols are used: p > .05.

The symbols are stating whether, based on the strength of the evidence, the probability of a Type I Error is less than (significant) or greater than (not significant) the generally accepted threshold of 5%.

Deeper Dive: What are Degrees of Freedom?

Degrees of Freedom (df) tell how much information you have that is free to vary. We are measuring things that can vary (i.e. variables) and are looking for specific patterns in how they vary. We use the information from a sample to estimate what is true in the population. Therefore, each case or person provides another piece of information we can use to estimate a parameter in the population.

When we estimate based on these variables, however, the variables are no longer fully independent. Each time we estimate a parameter using our data, the data are constrained. Each test that we use requires us to make at least one estimate about a population parameter using a sample statistic; each time we estimate a parameter we lose one degree of freedom.

When the constraints are groups (which is the case for *t*-tests), the number of independent groups is summarized with the symbol k. Degrees of freedom for each *t*-test are calculated as the total sample size minus the number of groups. This can be summarized as $n-k$. However, because each *t*-test has a different but stable number of groups, the df formula is often written with the specifics for k indicated as follows:

In the one sample *t*-test, there is only one sample so k = 1. Thus, our *df* for the one sample *t*-tests is calculated as $n-1$.

In the dependent samples *t*-test, there is only one sample being tested two times so k = 1. Thus, our df for the dependent samples *t*-test is calculated as $n-1$.

In the independent samples *t*-test we estimate two parameters: the mean for group 1 and the mean for group 2. Thus, our df is calculated as $n-2$. You may also see this specified by group so that df is calculated as n_1-1+n_2-1 . Each of these will yield the same number.

Reading Review 7.2

1. Which two things should be stated first in a results summary paragraph?
2. What are the four parts of the evidence string and what does each one report?
3. Which set of symbols should be used at the end of a sentence to indicate that a result was not significant?

4. What does reporting Cohen’s d add to a results summary?

7.5: Practice Example of how to Compute t when Descriptive Statistics are Provided

Usually, data are provided and used to compute the values needed for an inferential test such as the one sample *t*-tests. However, you may encounter times when the sample descriptive statistics are already available and do not need to be computed from the raw data. When this occurs, the preparatory steps of the analysis are already completed and it is possible to move to solving using the formula. Now that we have walked through all the preparatory steps and steps to solving in detail using sample data, let's practice one more time with just the steps for solving.

Let us assume we want to assess whether the students at one school called *Valley College* are scoring significantly higher on a standardized test than students in general in the United States. Let us assume that Valley College had 36 students whose mean score on the test was 80.00 with a standard deviation of 10.00. Let us also assume that we know the mean for the population of students in the U.S. to be 55.00.

Let's use this information to follow the steps in hypothesis testing:

1. State the hypothesis.
   - It is hypothesized that students at Valley College will have a higher mean test score than the mean of 55.00 for students in general in the United States.
2. Choose the inferential test (formula) that best fits the hypothesis.
   - A one sample *t*-test is the best fit for comparing the mean from a sample to the known mean of a population.
3. Determine the critical value.
   - The standard alpha level of .05 can be used. The hypothesis is directional, so a one-tailed test is required. The sample size is 36 so the *df* = 35. Thus, the critical value associated with a *df* of 35 in Appendix D should be used. Below is an excerpt of the section of the *t*-tables that best fits the current hypothesis and data. Under the conditions of an alpha level of .05, a one-tailed test, and 35 degrees of freedom the critical value is 1.690.

	one-tailed test	
alpha level:	α = 0.05	α = 0.01
Degrees of Freedom:	35	35
Critical Values:	1.690	2.438

4. Calculate the test statistic.

<u>Section A: Preparation</u>

The preparatory values are provided and only need to be identified rather than computed. For the current example for a one sample *t*-test, this preparatory values are as follows:

μ = 55.00

n = 36

$\bar{x}$= 80.00

s = 10.00

<u>Section B: Solving</u>

Now that the preparatory work is done, the formula can be used to compute the obtained value. For the one sample *t*-test, this work is as follows:

1. Write the formula with the values found in section A plugged into their respective locations.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$
$$t = \frac{80.00 - 55.00}{\frac{10.00}{\sqrt{36}}}$$
$$t = \frac{25.00}{\frac{10.00}{6.00}}$$
$$t = \frac{25.00}{1.666\ldots}$$
$$t = 15.00$$

5. Apply a decision rule and determine whether the result is significant.

   The critical value is 1.690

   The obtained *t*-value is 15.00

   The obtained *t*-value exceeds (i.e. is greater than) the critical value, thus, the result is strong enough to be significant.

   However, an additional consideration is needed when the hypothesis is directional. When a hypothesis is directional, the result must both be strong enough to be significant (by being greater in absolute value than the critical value) and also in the same direction as was hypothesized.

   It was hypothesized that the mean for Valley College would be higher than that of the United States in general. The mean for Valley College was 80.00 and the mean for the U.S. in general was 55.00. Thus, the data are in the same direction as hypothesized.

   In summary, because the data are in the direction hypothesized *and* the obtained value was greater than the critical value, the result can be determined to be significant.

6. Calculate the effect size.

   Cohen's d can be used to calculate the effect size for a one sample *t*-test. The Cohen's *d* calculations for the current data would be as follows:

$$d = \frac{\bar{x} - \mu}{s}$$
$$d = \frac{80.00 - 55.00}{10.00}$$
$$d = \frac{25.00}{10.00}$$
$$d = 2.50$$

   Using the rules of thumb, this would be considered a large effect size.

7. Report the results in American Psychological Associate (APA) format.

   A one sample *t*-test was used to test the hypothesis that students at Valley College would have a higher mean test score than the mean of 55.00 for students in general in the United States. Consistent with the hypothesis, the mean score for the sample (M = 80.00; SD = 10.00) was significantly higher than the mean for the population, $t(35) = 15.00$, $p < .05$. The Cohen's d effect size of 2.50 was large.

7.6: Using SPSS

As reviewed in Chapter 2, software such as SPSS can be used to expedite analyses once data have been properly entered into the program. Data need to be organized and entered into SPSS in ways that serve the analysis to be conducted. Thus, this section focuses on how to enter and analyze data for a one sample *t*-test using SPSS. SPSS version 29 was used for this book; if you are using a different version, you may see some variation from what is shown here.

Entering Data

The one sample *t*-test is univariate. The variable being tested must be quantitative and should have been measured using numbers on an interval or ratio scale. If this is true, you are ready to open SPSS and begin entering your data.

Open the SPSS software, click “New Dataset,” then click “Open” (or “OK” depending on which is shown in the version of the software you are using). This will create a new blank spreadsheet into which you can enter data. There are two tabs which appear towards the bottom of the spreadsheet. One is called “Variable View” which is the tab that allows you to tell the software about your variables. The other is called “Data View” which is the tab that allows you to enter your data.

Click on the Variable View tab. This tab of the spreadsheet has several columns to organize information about the variables. The first column is titled “Name.” Start here and follow these steps:

1. Click the first cell of that column and enter the name of your test variable using no spaces, special characters, or symbols. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
2. Click the first cell of the column titled “Type” and then click the three dots that appear in the right side of the cell. Specify that the data for that variable appear as numbers by selecting “Numeric.” For numeric data SPSS will automatically allow you to enter values that are up to 8 digits in length with decimals shown to the hundreds place as noted in the “Width” and “Decimal” column headers, respectively. You can edit these as needed to fit your data, though these settings will be appropriate for most quantitative variables in the behavioral sciences.
3. Click the first cell of the column titled “Label.” This is where you can specify what you want the variable to be called in output, including in tables and graphs. You can use spaces or phrases here, as desired.
4. Click on the first cell of the column titled “Measure.” A pulldown menu with three options will allow you to specify the scale of measurement for the variable. SPSS does not differentiate between interval and ratio data and, instead, refers to both of these as “Scale.” Select the “Scale” option because, if you are using a one sample *t*-test, your data should have been measured on the interval or ratio scale. If you are only planning to conduct one test, you can skip step 5 below.
5. If you will conduct additional one sample *t*-tests with other variables, move to the second row of the spreadsheet, starting with the cell under “Name” and repeat steps 1-4 until you have entered the information for all of your variables.

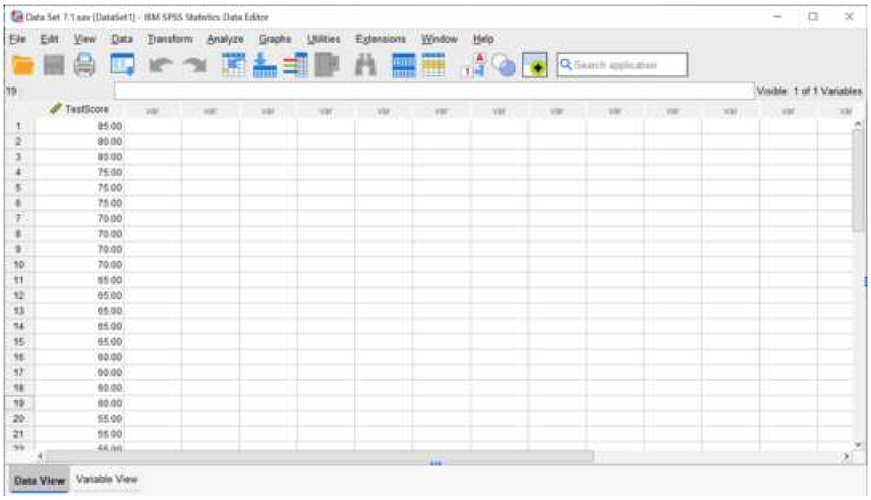
Here is what the Variable View tab would look like when created for Data Set 7.1:

Now you are ready to enter your data. Click on the Data View tab toward the bottom of the spreadsheet. This tab of the spreadsheet has several columns into which you can enter the data for each variable. Each column will show the names given to the variables that were entered previously using the Variable View tab. Click the first cell corresponding to the first row of the first column. Start here and follow these steps:

1. Enter the data for the test variable moving down the rows under the first column. If your data are already on your computer in a spreadsheet format such as excel, you can copy-paste the data in for the variable. Note: If the spreadsheet will not allow you to enter the information and/or makes a blunt tone when you try to enter the data, it means you have an error in the format of the data you are trying to enter or that they do not match the details you provided in the variable tab. Change the data format or go back to edit the information on the Variable View tab if this occurs, as appropriate. Then, return to the Data View tab to enter your data. If you are only entering and data for one variable and conducting one test, skip step 2 below.
2. Repeat step 1 for each variable until all of your data have been entered.
3. Then hit save to ensure your data set will be available for you in the future.

> **Note**
>
> Data entered into SPSS are saved in a file format that can only be opened in specific forms of software such as SPSS. Therefore, if you use a computer to make data files at school or work that has SPSS and try to open them on a different computer which does not have SPSS, you will not be able to. SPSS files can only be opened on devices which have access to SPSS software. Keep this in mind if you plan to use different devices while practicing the use of SPSS.

Here is what the Data View tab would look like for the first 21 cases when created for data set 7.1:

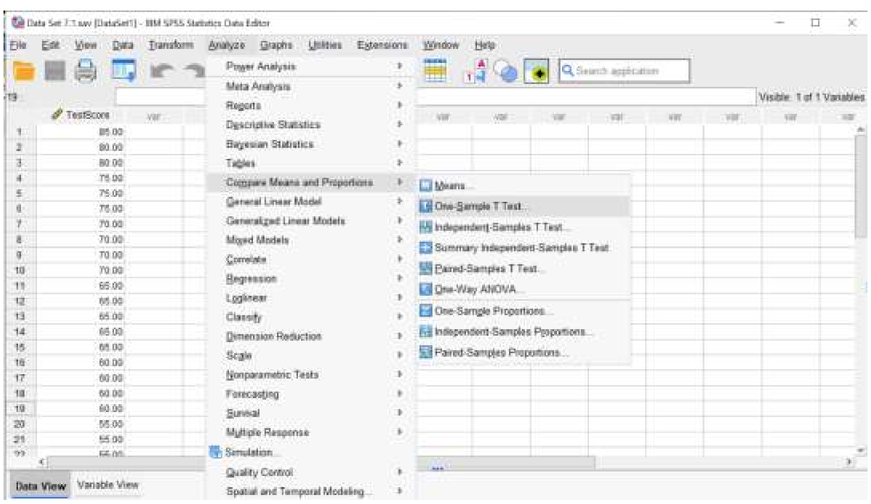


Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

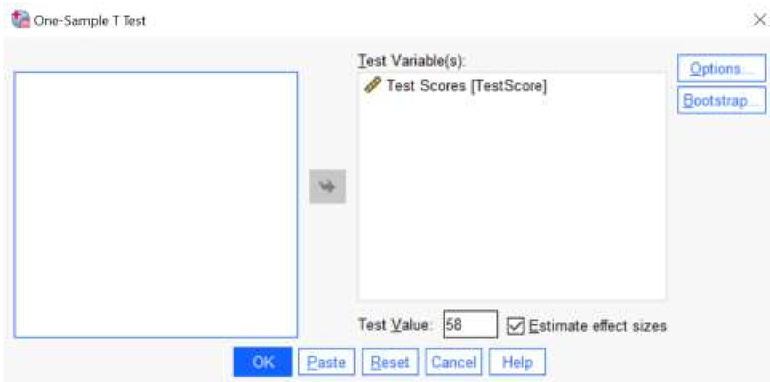
Conducting a One Sample t-Test in SPSS

The steps to running a one sample *t*-test in SPSS are:

1. Click Analyze › Compare Means and Proportions › One-Sample T Test from the pull-down menus as shown below.



2. Drag the name of the variable you want to test from the list on the left into the Test Variable box on the right of the command window. You can also do this by clicking on the variable name to highlight it and then clicking the arrow to move the variable from the left into the Variable text box on the right. Next, put the known or hypothesized population mean in the Test Value box under the Test Variable box. The population mean used for our example with Data Set 7.1 was 58.00 so this value is indicated in the Test Value box. If the version of SPSS you are using has a check box to estimate effect sizes (as shown in the picture below), click that as well.



3. Click OK.
4. The output (which means the page of calculated results) will appear in a new window of SPSS known as an output viewer. The results will appear in three tables as shown below.

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Test Scores	25	65.0000	10.20621	2.04124

One-Sample Test

	Test Value = 58						
			Significance			95% Confidence Interval of the Difference	
	t	df	One-Sided p	Two-Sided p	Mean Difference	Lower	Upper
Test Scores	3.429	24	.001	.002	7.00000	2.7871	11.2129

One-Sample Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
Test Scores	Cohen's d	10.20621	.686	.243	1.117
	Hedges' correction	10.53963	.664	.235	1.082

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Reading SPSS Output for a One Sample t-Test

The first table shows the descriptive statistics for the test. These include the sample size, mean, standard deviation, and standard error. These are foundational pieces that would appear in the formula steps as we saw when we performed hand-calculations for Data Set 7.1 earlier in this chapter.

The second table shows the main test results which are needed for the evidence string. These include the *t* value, the degrees of freedom (*df*) and the *p*-value for a one-tailed test (called "One-Sided p" in SPSS) and for a two-tailed test (called a Two-Sided p in SPSS). When using the standard alpha level of .05, a *p*-value that is less than .05 is significant. This is because the *p*-value shown in SPSS is the risk of a Type I Error (which is what the alpha level is saying needs to be less than .05). Our hypothesis for Data Set 7.1 earlier in this chapter was non directional and, thus, the two-tailed (i.e. two-sided) *p*-value should be used. SPSS reports this value to be .002. The obtained *p*-value of .002 is less than the alpha level of .05, thus, the result is significant. This conclusion is consistent with the conclusion we had when using hand-calculations to compare the obtained *t*-value to the critical value that fit our hypothesis and data. Therefore, the results and conclusions when using hand calculations and when reading the results of SPSS are consistent with one another. This is what should always happen unless a mistake was made in either the use of hand-calculations or SPSS.

Take a careful look at the results (i.e. the SPSS output) and you will see they match what we found when doing computations by hand with Data Set 7.1.

> **Note**
>
> You will sometimes see slight variation in results due to rounding error when comparing hand-calculated results to SPSS generated results. However, these differences should usually only appear in the third (thousandths) or fourth (ten-thousandths) decimal place; your hand calculated results will usually match the SPSS generated results to the hundredths place (meaning they should match to two decimal places).

The last table of results in the SPSS output shows the effect sizes. This will only appear if you checked the box in the command window to select this extra analysis. By default, SPSS will provide two calculations of effect size: Cohen's *d* and Hedge's correction. The results of these calculations will be similar for data sets with sample sizes of 20 or greater. When a sample size is approximately 20 or more, Cohen's *d* can be used. When the sample size is lower than 10, the Hedge's correction may be more appropriate. Our sample size was 25 and the results of both of these effect size estimates are similar, thus, the Cohen's d value is appropriate to use. The effect size is reported in the SPSS output table in the column labelled "point estimate." SPSS reports Cohen's *d* in this column as 0.686, or 0.69 when rounded to the hundredths place. If you compare the result for Cohen's *d* shown in the SPSS output table to our hand-calculations for earlier, you will see they are the same.

Reading Review 7.3

1. What scale of measurement should be indicated in SPSS for the test variable?
2. Under which table and column of the SPSS output can the t-value be found?
3. Under which table and column of the SPSS output can the p-value be found?
4. Under which table and column of the SPSS output can the d-value be found?

7.7: Structured Summary for the One Sample t-Test

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes.

Summarize what each symbol stands for.

μ =

$\bar{x}$ =

s =

n =

Fill-in the appropriate information for each section below:

1. One Sample *t*-Test Basics
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What assumptions must the data meet to use this test?
2. One Sample *t*-Test Formula
   a. What is the formula for a one sample *t*-test?
   b. What are the preparatory steps for using this formula?
   c. What are the steps for computing *t* using this formula?
3. Reporting Results from a One Sample *t*-Test
   a. How is this statistic reported when using APA format?
      i. What three things must be reported?
      ii. What are the parts of the evidence string and what does each stand for or indicate?

7.8: References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum

Navarro, D. J. (2014). Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.4). Self-published online.

CHAPTER OVERVIEW

8: Independent Samples t-Tests

The **independent samples *t*-test** is an inferential test used when you want to test whether the means of two unrelated groups are significantly different. This is sometimes referred to as a two sample *t*-test, a between samples *t*-test, an unpaired *t*-test, or a student *t*-test (that last name has an interesting origin). The independent samples *t*-test is a bivariate test used when there are two separate groups to compare. The two groups are compared on a single quantitative variable which is measured the same way for both groups. Thus, you should use this technique if you want to compare the means of exactly two independent groups on a variable measured on the interval or ratio scale. Other techniques are needed when there is one group (such as the one sample *t*-test covered in Chapter 7) or three or more groups (such as the one-way ANOVA which will be covered in Chapter 10).

8.1: Variables, Data, and Hypotheses that Fit the Independent Samples t-Test

The independent samples *t*-test is a bivariate test. This means two variables are measured and used. One of those variables must be a qualitative grouping variable with exactly two groups and the other must be a quantitative variable measured using an interval or ratio scale. An example of a qualitative grouping variable with two groups would be job status where participants are either grouped as being “Employed” or “Unemployed.” Another example would be tutoring attendance where participants are grouped as either “Have Attended” or “Have Not Attended.” The grouping variable is used to distinguish the two groups which are going to be compared. For simplicity, groups are often referred to as Group 1 and Group 2.

The thing being compared between the two groups is the quantitative variable. This is sometimes considered an outcome variable. An example of a quantitative variable would be hours of sleep. In this example, all participants would have their hours of sleep measured and used in the analysis. Another example of a quantitative variable would be exam scores. In this example, all participants would complete an exam and their scores on the exam would be computed and compared in the analysis. For example, if a researcher wanted to test whether those who attended tutoring had greater mean exam scores than those who did not, the qualitative grouping variable would be tutoring attendance and the quantitative variable being compared between the two groups would be exam scores.

Data

Each inferential test has some assumptions which must be met in order for the formula to function properly. In keeping, there are a few assumptions about the data which must be met before an independent samples *t*-test is used. First, the data for the quantitative variable must have been measured the same way for all cases in each of the two groups. You cannot use different ways of measuring the quantitative variable in each group; the test or measure used to obtain the scores to test cases in Group 1 should be the same as the measure used in Group 2. Second, the data for the quantitative variable must be measured on the interval or ratio scale of measurement. Third the members of the two groups must be non-overlapping and independent of one another. This means that no participant can be in both groups; each participant can only be a member of one of the two groups. You can see this important requirement reflected in most of the different names used for the independent samples *t*-test (i.e. “independent samples,” “two sample,” “unpaired”). Fourth, data for the quantitative variable should be fairly normally distributed in each group. Finally, there should be homogeneity of variances. **Homogeneity of variance** is when the variances for the quantitative variable are similar in both groups. When variances are not homogeneous it means that the two groups have different amounts of error (as estimated using variance) so their distribution curves have different widths and/or heights. When variances are not homogenous, adjustments to the formula are required. We will review that later in this chapter. For now, if all five of these assumptions are met, the independent samples *t*-test can be used.

Hypotheses

Hypotheses for the independent samples *t*-test must include both the qualitative variable and the quantitative variable and can be either non-directional or directional. For the independent samples *t*-test, the non-directional research hypothesis is that the sample means will be different from each other. The corresponding null hypothesis is that the sample means will not be different from each other. Because this research hypothesis is non-directional, it requires a two-tailed test. The non-directional research and corresponding null hypotheses can be summarized as follows:

Non-Directional Hypothesis for an Independent Samples t-Test

Research hypothesis	The mean of Group 1 is not equal to the mean of Group 2.	$H_A: \mu_1 \neq \mu_2$
Null hypothesis	The mean of Group 1 is equal to the mean of Group 2.	$H_0: \mu_1 = \mu_2$

There are two directional hypotheses possible for the independent samples *t*-test. One possible directional research hypothesis is that the mean for Group 1 will be *greater than* the mean for Group 2. The corresponding null hypothesis is that the mean for Group 1 will *not* be greater than the mean for Group 2. This could mean that the mean for Group 1 is less than or that it is equal to the mean for Group 2. Because this research hypothesis is directional, it requires a one-tailed test. This version of the research and corresponding null hypotheses can be summarized as follows:

Directional Hypothesis for an Independent Samples t-Test: Version 1

Research hypothesis	The mean of Group 1 will be greater than the mean of Group 2.	$H_A: \mu_1 > \mu_2$
Null hypothesis	The mean of Group 1 will be less than or equal to the mean of Group 2.	$H_0: \mu_1 \leq \mu_2$

For the independent samples *t*-test, the other possible directional research hypothesis is that the mean for Group 1 will be *less than* the mean for Group 2. The corresponding null hypothesis is that the mean for Group 1 will *not* be less than the mean for Group 2. This could mean that the mean for Group 1 is greater than or that it is equal to the mean for Group 2. Because this research hypothesis is directional, it requires a one-tailed test. This version of the research and corresponding null hypotheses can be summarized as follows:

Directional Hypothesis for an Independent Samples t-Test: Version 2

Research hypothesis	The mean of Group 1 will be less than the mean of Group 2.	$H_A: \mu_1 < \mu_2$
Null hypothesis	The mean of Group 1 will be greater than or equal to the mean of Group 2.	$H_0: \mu_1 \geq \mu_2$

These three version of the hypothesis are the broad form and would be refined to include the specific names or categories of the variables that are under investigation when working with specific hypotheses. For example, if a researcher expected that students who were given study guides would have different exam scores than those who were not given study guides, the research and null hypotheses would be written as follows:

Specific, Non-Directional Hypothesis for an Independent Samples t-Test

Research hypothesis	The mean exam score for those who received study guides will not be equal to the mean exam score for those who did not receive study guides.	$H_A: \mu_1 \neq \mu_2$
Null hypothesis	The exam score for those who received study guides will be equal to the mean exam score for those who did not receive study guides.	$H_0: \mu_1 = \mu_2$

However, it is possible to also propose a directional hypothesis with these two variable: study guides (as the qualitative, grouping variable) and exam scores (as the quantitative variable). For example, if the researcher expected that students who were given study guides would have exam scores that were higher than those who are were not given study guides, the research and null hypotheses would be directional and could be written as follows:

Specific, Directional Hypothesis for an Independent Samples t-Test

Research hypothesis	The mean exam score for those who received study guides will be greater than the mean exam score for those who did not receive study guides.	$H_A: \mu_1 > \mu_2$
Null hypothesis	The exam score for those who received study guides will be less than or equal to the mean exam score for those who did not receive study guides.	$H_0: \mu_1 \leq \mu_2$

8.2: Experimental Design and Cause-Effect

The independent samples *t*-test is sometimes used to analyze data from an experiment. Not all inferential tests fit well with experimental designs. This is because different designs yield different kinds of data and, thus, vary in which inferential tests best suit them. A misconception some folks have is that the words “research,” “study,” and “experiment” are synonymous and all refer to any work where data are collected and used to test hypotheses. However, experiments are specific types of studies. Study and research are broad terms whereas experiment is specific to only some forms of research. Thus, all experiments are studies or research but not all research or studies are experiments. Instead, experiments fall under a specific subcategory of research known broadly as “experimental designs.” Experiments are unique and especially valuable forms or research so it is worth taking a moment to understand them.

The **experimental design** of research involves controlling, manipulating, or constraining one variable to see if it has an impact on another variable. Because of this (and some additional features), experimental designs are the only kinds of studies in which cause-effect relationships can be deduced. The additional considerations when using an experimental design are beyond the scope of this book so we will stay focused on the key feature: Controlling the independent variable. Recall (from Chapter 1) that in a cause-effect relationship, the causal variable is called the independent variable (IV) and the affected variable is called the dependent variable (DV). In order to see if the IV impacts the DV in an experiment, researchers control how participants in each group experience the IV and then measure the DV. Control in this context means the researcher gets to decide what happens for the IV for each case. A common form of experiment is the experimental vs. control group design. In this design there are two groups (as indicated in the name of the design): an experimental group and a control group. Participants are often randomly assigned to one of these two groups. What distinguishes these two groups is whether (or to what extent) they experience the IV. The **experimental group** receives/is exposed to the IV (or a higher amount of the IV) whereas the **control group** does not receive/is not exposed to the IV (or gets a markedly lower amount of the IV). A control group provides a baseline for comparison. If the IV impacts the DV, then the experimental and control groups should have different levels of the DV as a result of their different experiences with the IV.

Let’s walk through these basic components of an experiment using the hypothesis that those who receive study guides will have higher exam scores than those who do not. The IV would be the study guide. An experiment could be done where a group of students are randomly assigned to either receive a study guide or not. The DV would be exam scores. To gather data on the DV, participants in both the experimental and control groups could be asked to complete the same exam. The exam scores would then be used to compute the mean exam scores for each of the two groups. If the mean for the experimental group (those who received study guides) is significantly greater than the mean for the control group (those who did not receive study guides), the directional hypothesis would be supported. In order to test whether the means are significantly greater in one group than the other, an independent samples *t*-test must be used. This is just one of many examples of times when data between two different groups could be tested using an independent samples *t*-test.

Non-Experimental Designs

The independent samples *t*-test is flexible and can also be used to compare the means of two different groups when data were acquired using a non-experimental research design. Therefore, though it is often used to test data from experiments by comparing an experimental group to a control group, it can also be used to test data from pre-existing groups in non experimental research. For example, the independent samples *t*-test can be used to compare the sleep hours of teens to those of young adults. It could also be used to compare the mean commute times of people who work morning shifts to those who work night shifts or to compare the satisfaction of clients who worked with either of two different salespersons.

When to Use Causal Language

Because the independent samples *t*-test can be used with experimental designs that deduce cause-effect and with non-experimental designs which cannot be used to deduce cause-effect, it is necessary to use causal language only when appropriate. Causal language includes words and phrases such as “caused, “impacted,” or “resulted in” to connect the variables. For example, if we say “Study guides caused mean exam scores to differ” or “The availability of study guides resulted in higher exam scores” we are making a claim that a causal relationship exists. Causal language can also be subtler. For example, if we say “Study guides led to differences in scores,” “Study guides made scores higher,” or even “Those who get study guides will get higher scores, on average” we are making causal claims. This kind of language should only be used when an experiment has been performed, not simply anytime two groups are compared and/or when independent samples *t*-tests are used. Therefore, it is best to use non causal language as a default and to only switch to using causal language when it is known that an experimental design was used and that causal language is appropriate.

Here are examples of how the claims from the prior paragraph could be rewritten without using causal language: Differences in scores were observed for those who did and did not receive study guides; The mean exam score was higher for the group which received study guides than for the group which did not receive study guides; Those who got study guides had higher scores, on average, than those who did not.

Reading Review 8.1

1. What is the general research hypothesis that can be tested using a one-tailed, independent samples *t*-test where Group 1 is expected to have the lower mean? Provide both sentence and symbol formats.
2. What is meant by *homogeneity of variances*?
3. How many grouping variables can be used each time an independent samples *t*-test is performed?
4. Under what conditions can causal language be used when performing or interpreting results using the independent samples *t*-test?

8.3: The Independent Samples t-Test Formula

Before we can use the formula, it is important to understand what it can tell us and how it gets there. The independent samples *t*-test formula looks quite complicated but is meant to tell us something that is conceptually fairly simple: The obtained *t*-value tells us far apart the two group means are using standard error. Another way to say this is that it tells us how many standard errors apart the two group means are. It does this by taking the difference in the sample means and dividing it by the standard error. The difference in means is calculated in the numerator of the formula and the pooled standard error is calculated in the denominator of the formula. The standard error (SE) used is known as a pooled standard error because it is calculated by putting information from both groups together (also known as *pooling information*). Therefore, we can understand the formulas main construction and outcomes as follows:

$$t=\frac{\text{difference in sample means}}{\text{pooled standard error}}=\text{how many SEs apart the two sample means are}$$

The numerator focuses on the difference between the mean of Group 1 and the mean of Group 2. This is the core of the formula because the hypotheses tested using an independent samples *t*-test are specifically asking whether these two means differ. Therefore, you can think of the numerator as the main focus of the formula and the denominator as taking into account other necessary information and adjustments. This will be true for the main format used for all three versions of the *t*-test in this book. The denominator of the independent samples *t*-test formula is used to take into account the error in the samples (via the pooled standard error).

The formula contains six symbols which represent and, thus, must be replaced with specific values. The independent samples *t*-test formula is as follows:

$$t=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\left[\frac{(n_1-1)\,s_1^2+(n_2-1)\,s_2^2}{n_1+n_2-2}\right]\left[\frac{n_1+n_2}{n_1\times n_2}\right]}}$$

Note

The df is the sum of $n–1$ for each group (or total n minus 2) and appears in the formula's denominator.

The numerator asks for the mean of Group 1 and the mean of Group 2. These will be used to calculate the difference in means. The denominator asks for the sample size for Group 1, the sample size for Group 2, the variance for Group 1, and the variance for Group 2. These four things will be used to calculate the pooled standard error. Thus, the formula requires we know three basic things about each of the two groups so that we can plug them in and solve: mean, variance, and sample size.

Though the formula looks very complex, focusing on a few things can help us see that it is manageable. First, though there appear to be many symbols in the formula there are only six and the simplest one of them, sample size, appears the most often. Thus, the majority of the formula is just asking us to input sample sizes. Second, the mathematical operations required are ones most of us are familiar with. Operations refer to mathematical actions we take. The operations this formula requires include: adding, subtracting, multiplying, dividing, squaring (if converting SD to variance), and square rooting. Though the formula requires us to do many steps, each step only includes one of these six basic operations (and we can use calculators to make the more challenging steps of squaring and square rooting easier).

Notice that the formula includes all three components that impact statistical power (see Chapter 6 for a review of the components of power):

1. The size of the change, difference, or pattern observed in the sample
2. The sample size
3. The size of the error in the sample statistic(s)

The first component, size of difference, is being incorporated in the numerator. The denominator is a bit more complex and it can be hard to see what it is doing so let's take a moment to consider its role. The denominator allows necessary adjustments to be made that incorporate the other two components of statistical power: sample size and error. Error is being measured using the variance (which is the squared version of standard deviation) of each group. These components are both incorporated into the formula. The way sample sizes and variances are put together in the denominator is used to find the pooled standard error. Pooled standard error, therefore, takes into account two of the three components of power: the size of the sample and the size of the error.

Let's take a moment to see the connection between standard error for one group (which we reviewed in Chapter 6) and standard error when it is pooled so that it takes into account two groups. Standard error for a single sample is found using this formula:

$$SE=\frac{\sigma}{\sqrt{n}}$$

The independent samples *t*-test needs to pool this from two separate groups. It does so by first finding the pooled standard deviation (because there are two groups so there are two standard deviations) and adjusting that by a calculation of samples sizes. The left section of the denominator is calculating the pooled standard deviation. The **pooled standard deviation** (S_{pooled} or S_p) is the weighted average of the standard deviation of two or more groups. When we say an average is "weighted" it means that proportionally more "weight" is given to groups with larger sample sizes. Another way of saying this is that a weighted average takes into account any differences in sample size so that the group(s) with more cases (i.e. larger sample sizes) more heavily impact the value being calculated. This is how pooled standard deviation can be found when group variances are approximately even:

$$S_p=\sqrt{\left[\frac{\left(n_1-1\right)s_1^2+\left(n_2-1\right)s_2^2}{n_1+n_2-2}\right]}$$

You may wonder why pooled standard deviation is calculated using variances and sample sizes rather than standard deviations and sample sizes. Recall from Chapter 4 that variance is used when we actually want the standard deviation but still have some work to do before we do the final step of square rooting. Notice that the S_p formula asks us to do calculations with variances and sample sizes but that those are all under a square root sign. Thus, we are using variance in route to finding the pooled standard deviation (which we will get to by square rooting as the last step of the S_p formula).

This needs to be further adjusted to go from pooled standard deviation to pooled standard error. Within the brackets on the right side of the denominator, we see the use of just the sample sizes. That piece is also under the square root sign so, on its own, it looks like this:

$$\sqrt{\left[\frac{n_1+n_2}{n_1\times n_2}\right]}$$

The presence of this piece changes the denominator of the formula from just finding the S_p into a formula for finding pooled standard error (SE_p). It is doing the same work that is done by the denominator of the SE formula for one group. Therefore, the formula for pooled standard error for an independent samples *t*-test looks like this:

$$SE_p=\left(\sqrt{\left[\frac{\left(n_1-1\right)s_1^2+\left(n_2-1\right)s_2^2}{n_1+n_2-2}\right]}\right)\left(\sqrt{\left[\frac{n_1+n_2}{n_1\times n_2}\right]}\right)$$

This can rewritten for (slightly more) simplicity (though, admittedly, it still looks a bit complex) as follows:

$$SE_p=\sqrt{\left[\frac{\left(n_1-1\right)s_1^2+\left(n_2-1\right)s_2^2}{n_1+n_2-2}\right]\left[\frac{n_1+n_2}{n_1\times n_2}\right]}$$

Therefore, the denominator of the independent samples *t*-test formula is the formula for pooled standard error for the test.

Let's put the numerator and denominator together. The numerator is used to find the difference in the means. The denominator is used to find the relevant standard error for the formula. Finally, the formula tells us to divide the numerator by the denominator. When we do, we get a *t*-value which tells us how many standard errors the two means are apart.

Thus, we have come full circle and can now (hopefully) see why the formula can be thought of as follows:

$$t=\text{how many standard errors apart the two sample means are}$$

Interpreting Obtained t-Values

Obtained *t*-values have two components: a magnitude and a direction. The magnitude is the absolute value of *t* and it represents how many standard errors the mean of one sample is from the mean of the other sample. The larger the value, the farther apart the two means are. As the *t*-value increases, the evidence for the research hypothesis and against the null hypothesis also increases. Conversely, as the *t*-value decreases, the evidence for the research hypothesis and against the null hypothesis also decreases. Thus, researchers are generally hoping for larger *t* values. The other component of *t* is its direction. When *t* is positive, it indicates that Group 1 had the higher mean than Group 2. Conversely, when *t* is negative, it indicates that Group 1 had the lower mean than Group 2. If the obtained *t*-value was 0.00, it would indicate that the means were zero standard errors apart (which is the same as saying the means were equal). If, for example, the obtained *t*-value was 3.00 it would indicate that the means of the two groups were three standard errors apart and that Group 1 had the higher mean. However, if the obtained *t*-value was -3.00 it would indicate that the two means were three standard errors apart and that Group 2 had the higher mean.

When testing a two-tailed (non-directional) hypothesis, only the magnitude needs to be considered to determine whether a result is significant. This is because a two-tailed hypothesis will be significantly supported if the difference in the means is sufficiently large regardless of which group mean was higher. However, when testing a one-tailed (directional) hypothesis, both magnitude and direction need to be considered. When it is hypothesized that Group 1 will have the higher mean than Group 2, the hypothesis will be significantly supported if the difference in the means is sufficiently large and the result is positive. When it is hypothesized that Group 2 will have the higher mean than Group 1, the hypothesis will be significantly supported if the difference in the means is sufficiently large and the result is negative. Thus, the direction of the results must match the direction of the hypothesis when using a one-tailed test of significance.

Alternative Ways to Write the Independent Samples t-Test Formula.

You may see the independent samples *t*-test written other ways. One common version is this:

$$t=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\left[\frac{SS_1+SS_2}{n_1+n_2-2}\right]\left[\frac{n_1+n_2}{n_1\times n_2}\right]}}$$

Another common version is this:

$$t=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\left[\frac{SS_1+SS_2}{n_1+n_2-2}\right]\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$$

These formulas are just different ways of getting to the same end result which some people find more or less intuitive. Let's take a quick look at the variations in these formulas and how they are equivalent to what is in the version of the formula shown earlier. First, both of these versions use sum of squares (SS) in place of adjusted sample size multiplied by variance for each group. This is because variance is calculated by dividing SS by adjusted sample size (calculated as $n-1$ when working with samples). Thus, if we multiple variance by $n-1$ it turns it back into SS.

$$\text{Variance Formula: } s^2=\frac{\Sigma(X-\bar{X})^2}{n-1}$$

$$\text{Sum of Squares Formula: } SS=\Sigma(X-\bar{X})^2$$

We can write it out to demonstrate. Variance times adjusted sample size is equal to SS because the adjusted sample sizes cancel out like so:

$$s^2=\frac{\Sigma(\mathrm{X}-\bar{\mathrm{X}})^2}{n-1}$$

$$s^2(n-1)=\frac{\Sigma(\mathrm{X}-\bar{\mathrm{X}})^2}{\cancel{n-1}}\cancel{(n-1)}$$

$$SS=\Sigma(\mathrm{X}-\bar{\mathrm{X}})^2$$

For this reason, SS can be used in place of $(n-1)s^2$ for each group in the denominator of the formula.

The other option for rewriting the formula is to use $\left[\frac{1}{n_1}+\frac{1}{n_2}\right]$ in place of $\left[\frac{n_1+n_2}{n_1 \times n_2}\right]$ in the right side of the denominator of the formula. These are just two way one can write the same things.

$$\left[\frac{1}{n_1}+\frac{1}{n_2}\right]=\left[\frac{n_1+n_2}{n_1 \times n_2}\right]$$

Therefore, though we have seen three different ways to write the independent samples *t*-test formula, they are all simply different ways or writing the same mathematical concepts and steps and, thus, will all yield the same result.

Reading Review 8.2

1. What is being calculated and represented by the numerator of the independent samples *t* test formula?
2. What is being calculated and represented by the denominator of the independent samples *t*-test formula?
3. Without making a determination of significance, how might t = 0.00 be interpreted?
4. Without making a determination of significance, how might t = 1.46 be interpreted?
5. Without making a determination of significance, how might t = -3.75 be interpreted?

Formula Components

Now that we have taken some time to understand the construction of the independent samples *t*-test formula, let's focus on how to actually use it, starting with identifying all of its parts.

In order to solve for *t*, six things must first be known:

$\bar{X}_1$ = the mean for Group 1

$\bar{X}_2$ = the mean for Group 2

S_1^2 = the variance for Group 1

S_2^2 = the variance for Group 2

n_1 = the sample size for Group 1

n_2 = the sample size for Group 2

We use the raw scores from each group to calculate their respective means, variances, and sample sizes.

Formula Steps

The steps are shown in order and categorized into two sections: A) preparation and B) solving. I recommend using this categorization to help you organize, learn, and properly use all inferential formulas. Preparation steps refer to any calculations that need to be done before values can be plugged into the formula. For the independent samples *t*-test this includes finding the descriptive statistics that make up the six components of the formula for each group (listed in the section above). Once those are known, the steps in section B can be used to yield the obtained value for the formula. The symbol for the obtained value for each *t*-test is *t*. Follow these steps, in order, to find *t*.

Section A: Preparation

1. Find n for Group 1.
2. Find $\bar{x}$ for Group 1.
3. Find s^2 for Group 1.
4. Find n for Group 2.
5. Find $\bar{x}$ for Group 2.
6. Find s^2 for Group 2.

Section B: Solving

1. Write the formula with the values found in section A plugged into their respective locations.
2. Solve the numerator by subtracting the mean of Group 2 from the mean of Group 1.
3. Solve for the left side of the denominator as follows:

a. Multiply variance for Group 1 by $n-1$ for Group 1 to get the SS for Group 1.
b. Multiply variance for Group 2 by $n-1$ for Group 2 to get the SS for Group 2.
c. Add the SS for Group 1 (which is the result of step 3a) to the SS for Group 2 (which is the result of step 3b). This gives you total SS.
d. Find the degrees of freedom (df) by adding the sample size for Group 1 to the sample size for Group 2 and then subtracting 2 from that total.
e. Divide the total SS (which is result of step 3c) by the df (which is the result of step 3d).

4. Solve right side of the denominator as follows:
   1. Add the sample size of Group 1 to the sample size of Group 2 to get the total sample size.
   2. Multiply the sample size of Group 1 by the sample size of Group 2.
   3. Divide the total sample size (which is the result of step 4a) by the product of sample sizes (which is the result of step 4b).
5. Multiply the left side of the denominator (which is the result of step 3e) by the right side of the denominator (which is the result of step 4c).
6. Square root the denominator (which means square root the result of step 5) to get the pooled standard error for the formula.
7. Finally, divide the numerator (which is the result of step 2) by the pooled standard error (which is the result of step 6) to get the obtained t-value.

8.4: Example of how to Test a Hypothesis by Computing t.

Let us assume that a researcher believed that high school students would have a higher mean for hours of sleep than college students. Suppose that sleep was measured in hours for a group of 5 high school students and a separate group of 5 college students. Assume that Data Set 8.1 includes data from these two independent samples. Let's use this information to follow the steps in hypothesis testing.

Data Set 8.1. Hours of Sleep Reported by High School and College Students (n = 10)

High Schoolers	College Students
5	3
7	5
6	4
8	6
9	7

Steps in Hypothesis Testing

In order to test a hypothesis, follow these steps:

1. **State the hypothesis.**

A directional hypothesis is the best fit because the goal is to see if the high school students have a *higher* mean score that the college students. A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that the mean hours of sleep will be higher among high schooler students than college students.*

Directional Hypothesis for an Independent Samples t-Test

Research hypothesis	The mean hours of sleep for high school students will be greater than the mean for college students.	$H_A : \mu_1 > \mu_2$
Null hypothesis	The mean hours of sleep for high school students will not be greater (i.e. will be less than or equal to) the mean for college students.	$H_0 : \mu_1 \leq \mu_2$

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The means of two independent samples are being compared so the appropriate test is an independent samples *t*-test.

3. **Determine the critical value.**

In order to determine the critical value, three things must be identified:

i. the alpha level,
ii. whether the hypothesis requires a one-tailed test or a two-tailed test, and,
iii. the degrees of freedom (df).

The alpha level is often set at .05 unless there is reason to adjust it such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The hypothesis is directional so a one-tailed test should be used.

The df must also be calculated. Each inferential test has a unique formula for calculating df. In short, the formula for df for the independent samples t-test is as follows: $n_1 + n_2 - 2$. It appears in the lower left side of the independent samples t-test formula. The sample size for Group 1 (the high school students) was $5 (n_1 = 5)$. The sample size for Group 2 (the college students) was also $5 (n_2 = 5)$. Thus, df = 5 + 5 – 2 so the df for this scenario is 8.

These three pieces of information are used to locate the critical value from the test. The full tables of the critical values for *t*-tests are located in Appendix D. Below is an excerpt of the section of the *t*-tables that fits the current hypothesis and data. Under the conditions of and alpha level of .05, a one-tailed test, and 8 degrees of freedom, the critical value is 1.860.

Critical Values Table

	one-tailed test	
alpha level:	α = 0.05	α = 0.01
Degrees of Freedom: 8	1.860	2.896

The critical value represents the absolute value which must be exceeded in order to declare a result significant. It represents the threshold of evidence needed to be confident a hypothesis is true. The obtained value (which is called *t* in a *t*-test) is the amount of evidence present. When using a two-tailed test, only the absolute value of the critical value must be considered. However, the current hypothesis requires a one tailed test because the hypothesis is directional. Therefore, both the magnitude and direction of the obtained *t*-value must be considered. Because the hypothesis states that Group 1 will have the higher mean than Group 2, the *t*-value must be positive in addition to exceeding the magnitude of the critical value. Thus, in order for the result to significantly support the hypothesis is needs to be positive and exceed the critical value of 1.860.

Degrees of Freedom for an Independent Samples t-Test

Degrees of Freedom (df) indicate how much information you have that is free to vary. The degrees of freedom is equal to the total of each sample size minus 1 per group. It appears in the independent samples *t*-test like this:

$$df = n_1 + n_2 - 2$$

Thus, in an independent samples *t*-test, the degrees of freedom are calculated as the total number of cases minus the number of groups (because we subtract one from each group). Total number of cases can be summarized with the symbol N. N is calculated as follows:

$$N = n_1 + n_2$$

The number of independent groups is summarized with the symbol *k*. Degrees of freedom for each *t*-test is calculated as the total sample size minus the number of groups. This can be summarizes as follows:

$$df = N - k$$

4. **Calculate the test statistic.**

A test statistic can also be referred to as an obtained value. The formula needed to find the test statistic known as *t* for this scenario is as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}\right]\left[\frac{n_1 + n_2}{n_1 \times n_2}\right]}}$$

Section A: Preparation

Start each inferential formula by identifying and solving for the pieces that must go into the formula. For the independent samples *t*-test, this preparatory work is as follows:

1. Find n for Group 1.
   - This value is found using Data Set 8.1 and is summarized as $n_1 = 5$
2. Find $\bar{x}$ for Group 1.
   - This value is found using Data Set 8.1 and is summarized as $\bar{X}_1 = 7.00$
3. Find s^2 for Group 1.
   - This value is found using Data Set 8.1 and is summarized as $s_1^2 = 2.50$

4. Find n for Group 2.
   - This value is found using Data Set 8.1 and is summarized as $n_2 = 5$
5. Find $\bar{x}$ for Group 2.
   - This value is found using Data Set 8.1 and is summarized as $\bar{X}_2 = 7.00$
6. Find s^2 for Group 2.
   - This value is found using Data Set 8.1 and is summarized as $s_2^2 = 2.50$

> **Note**
>
> The summarized calculations are shown below for reference. For a detailed review of how to calculate a sample mean, see Chapter 3. For a detailed review of how to calculate a sample variance, see Chapter 4.

Data Set 8.1. Descriptive Statistics for Hours of Sleep Reported by High School and College Students (n = 10)

High Schoolers	$X - \bar{X}_1$	$(X - \bar{X}_1)^2$	College Students	$(X - \bar{X}_2)^2$	$(X - \bar{X}_2)^2$
5	(-2)	4	3	(-2)	4
7	0	0	5	0	0
6	(-1)	1	4	(-1)	1
8	1	1	6	1	1
9	2	4	7	2	4
$n_1 = 5$		$SS_1 = 10$	$n_2 = 5$		$SS_1 = 10$
$\boldsymbol{\bar{X}_1 = 7.00}$		$s_1^2 = 2.50$	$\boldsymbol{\bar{X}_2 = 5.00}$		$s_2^2 = 2.50$

Now that the pieces needed for the formula have been found, we can move to Section B.

Section B: Solving

The inferential formula is used to compute the obtained value. For the independent samples *t*-test, this work is as follows:

1. Write the formula with the values found in section A plugged into their respective locations.

Writing the formula in symbol format before filling it in with the values can help you recognize and memorize it. Here is the formula with the symbols:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}\right]\left[\frac{n_1 + n_2}{n_1 \times n_2}\right]}}$$

Here is the formula with values filled into their appropriate locations in place of their symbols:

$$t = \frac{7.00 - 5.00}{\sqrt{\left[\frac{(5-1)2.50 + (5-1)2.50}{5+5-2}\right]\left[\frac{5+5}{5 \times 5}\right]}}$$

2. Solve the numerator by subtracting the mean of Group 2 from the mean of Group 1. Note: Steps will appear in bold to show when they have been computed.

$$t = \frac{\mathbf{2.00}}{\sqrt{\left[\frac{(5-1)2.50 + (5-1)2.50}{5+5-2}\right]\left[\frac{5+5}{5 \times 5}\right]}}$$

3. Solve for the left side of the denominator as follows:

a. Multiply variance for Group 1 by $n-1$ for Group 1 to get the SS for Group 1.

$$t=\frac{2.00}{\sqrt{\left[\frac{\mathbf{10.00}+(5-1)2.50}{5+5-2}\right]\left[\frac{5+5}{5\times5}\right]}}$$

b. Multiply variance for Group 2 by $n-1$ for Group 2 to get the SS for Group 2.

$$t=\frac{2.00}{\sqrt{\left[\frac{10.00+\mathbf{10.00}}{5+5-2}\right]\left[\frac{5+5}{5\times5}\right]}}$$

c. Add the SS for Group 1 (which is the result of step 3a) to the SS for Group 2 (which is the result of step 3b). This gives you total SS.

$$t=\frac{2.00}{\sqrt{\left[\frac{\mathbf{20.00}}{5+5-2}\right]\left[\frac{5+5}{5\times5}\right]}}$$

d. Find the degrees of freedom (df) by adding the sample size for Group 1 to the sample size for Group 2 and then subtracting 2 from that total.

$$t=\frac{2.00}{\sqrt{\left[\frac{20.00}{\mathbf{8}}\right]\left[\frac{5+5}{5\times5}\right]}}$$

e. Divide the total SS (which is result of step 3c) by the df (which is the result of step 3d).

$$t=\frac{2.00}{\sqrt{[\mathbf{2.50}]\left[\frac{5+5}{5\times5}\right]}}$$

4. Solve right side of the denominator as follows:

a. Add the sample size of Group 1 to the sample size of Group 2 to get the total sample size.

$$t=\frac{2.00}{\sqrt{[2.50]\left[\frac{\mathbf{10}}{5\times5}\right]}}$$

b. Multiply the sample size of Group 1 by the sample size of Group 2 to get the product of the sample sizes.

$$t=\frac{2.00}{\sqrt{[2.50]\left[\frac{10}{\mathbf{25}}\right]}}$$

c. Divide the total sample size (which is the result of step 4a) by the product of sample sizes (which is the result of step 4b).

$$t=\frac{2.00}{\sqrt{[2.50][\mathbf{0.40}]}}$$

5. Multiply the left side of the denominator (which is the result of step 3e) by the right side of the denominator (which is the result of step 4c).

$$t=\frac{2.00}{\sqrt{[\mathbf{1.00}]}}$$

6. Square root the denominator (which means square root the result of step 5) to get the pooled standard error for the formula.

$$t = \frac{2.00}{\mathbf{1.00}}$$

7. Finally, divide the numerator (which is the result of step 2) by the pooled standard error (which is the result of step 6) to get the obtained *t*-value.

$$t = 2.00$$

This result, known as a test statistic or *t*-value, can also be referred to by the general term "obtained value." This result is positive meaning Group 1 (the high schoolers) had a higher mean than Group 2 (the college students). The means of these two groups were two standard errors apart as indicated by the magnitude of the result.

5. **Apply a decision rule and determine whether the result is significant.**

   Assess whether the obtained value for t exceeds the critical value as follows:

   The critical value is 1.860.

   The obtained *t* value is 2.00

   The obtained *t* value does exceed (i.e. is greater than) the critical value. However, because this is a one-tailed test (due to the directional hypothesis), we must also check the direction of the result during this step. The hypothesis stated that Group 1 (high schoolers) would have the higher mean so the result must be positive to support this hypothesis. Thankfully, it is. Therefore, both criteria are met to declare the result significant:

i. The magnitude of the obtained value is greater than the critical value and
ii. The result is in the hypothesized direction.

   Thus, the result is significant and supports the hypothesis

> **Note**
>
> If the hypothesis had been non-directional (and, thus, a two-tailed test was being performed), we would only need to check that the magnitude exceeded that of the critical value to declare the result significant.

6. **Calculate the effect size and/or other relevant secondary analyses.**

   When it is determined that the result is significant, effect sizes should be computed. Because the result was determined to be significant in step 5, the effect size is needed before proceeding to step 7 to complete the process.

   The effect size that is appropriate for *t*-tests under standard conditions is known as Cohen's d (Cohen, 1988). The formula for Cohen's d is as follows when working with an independent samples *t*-test:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{S_p}$$

   Cohen's d, when used for an independent samples *t*-test, calculates how many pooled standard deviations the sample mean of Group 1 is from the mean of Group 2. Thus, the numerator finds the difference in the two means and the denominator is used to divide that by the pooled standard deviation (S_p). Notice that this is very similar to the format of the *t*-test formula only that standard deviation is being used in place of standard error.

   Recall, that the S_p is in the boom of the denominator of the *t*-formula and can be isolated and used to compute S_p for Data Set 8.1 as follows:

$$S_p = \sqrt{\left[\frac{(n_1-1)\,s_1^2+(n_2-1)\,s_2^2}{n_1+n_2-2}\right]}$$
$$S_p = \sqrt{\left[\frac{[(5-1)2.50+(5-1)2.50}{5+5-2}\right]}$$
$$S_p = \sqrt{\left[\frac{10.00+10.00}{8}\right]}$$
$$S_p = \sqrt{\left[\frac{20.00}{8}\right]}$$
$$S_p = \sqrt{2.50}$$
$$S_p = 1.5811\ldots$$

The calculations for Cohen's d for Data Set 8.1 are as follows:

$$d = \frac{7.00-5.00}{1.5811\ldots}$$
$$d = \frac{2.00}{1.5811\ldots}$$
$$d \approx 1.2649$$

Effect sizes, like most values, are rounded and reported to the hundredths place. Thus, this effect size is reported as $d = 1.26$. Cohen's d can be interpreted using the following rules of thumb (Cohen, 1988; Navarro, 2014):

Interpreting Cohen's d Effect Sizes

~0.80	Large effect
~0.50	Moderate effect
~0.20	Small effect

Note that the magnitude of the d value is what is used to interpret effect, not the direction; thus, the absolute value of d is compared to the rules of thumb to estimate the effect size of the result.

The rules of thumb are general guidance and do not dictate precise or required interpretations. Instead, they provide some generally agreed upon approximations to aid in interpretations. As is true of all analyses, it is best to consider their situated (or practical) relevance. Nevertheless, the rules of thumb are useful in providing an initial guideline for interpreting effect sizes. Following these rules of thumb, the current finding of d = 1.26 would be considered a large effect.

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

Keep in mind that results are reported in past tense because they report on what has already been found. In addition, the research hypothesis must be stated but the null hypothesis is usually not needed for summary paragraphs because it can be deduced

from the research hypothesis. Finally, APA format requires a specific format be used for reporting the results of a test. This includes a specific format for reporting relevant symbols and details for the formula and data used.

Note: Standard deviations are reported with means, not variances. Thus, the standard deviation must also be computed before completing the write-up. Standard deviation for each group can be computed by square rooting their respective variances like so:

High School Group: $s_1^2 = 2.50$ Thus, $\sqrt{s_1^2} = \sqrt{2.50}$ so s = 1.5811… which rounds to 1.58

College Group: $s_2^2 = 2.50$ Thus, $\sqrt{s_2^2} = \sqrt{2.50}$ so s = 1.5811… which rounds to 1.58

Following this, the results for our hypothesis with Data Set 8.1 can be written as shown in the summary example.

APA Formatted Summary Example

An independent samples *t*-test was used to test the hypothesis that the mean hours of sleep would be higher among high schooler students than college students. Consistent with the hypothesis, the mean hours of sleep was significantly higher for the high schooler students (M = 7.00; SD = 1.58) than for the college students (M = 5.00; SD = 1.58), t(8) = 2.00, $p < .05$. The Cohen's d effect size of 1.26 was large.

This succinct summary in APA format provides a lot of detail and uses specific symbols in a particular order. To understand how to read and create a summary like this, see the brief review of the structure for APA format below.

Summary of APA-Formatted Results for the Independent Samples t-Test

In your APA-formatted write up for an independent samples *t*-test you should state:

1. Which test was used and the hypothesis which warranted its use.
2. Whether the aforementioned hypothesis was supported or not. To do so properly, three components must be reported:
   a. The mean and standard deviation for each group
   b. The test results in an evidence string as follows: $t(df)$ = obtained value
   c. The significance part of the evidence string as $p < .05$ if significant or $p > .05$, *ns* if not significant
   d. The effect size, if the result was significant

The following breaks down what each part represents in the evidence string for Data Set 8.1:

Anatomy of an Evidence String

Symbol for the test	Degrees of Freedom	Obtained Value	*p*-Value
t	(8)	= 2.00,	$p < .05$

Reading Review 8.3

1. Which two things should be stated first in a results summary paragraph?
2. What are the four parts of the evidence string and what does each one report?
3. Which set of symbols should be used at the end of a sentence to indicate that a result was significant?
4. What does reporting Cohen's d add to a results summary?

8.5: Addressing Violations to Assumptions

As noted earlier in this chapter, there are several assumptions to the independent samples *t* test. When they are met, we can proceed using the standard procedures described throughout this chapter. However, when assumptions of tests are not met it means either the test cannot be used or that a modified or alternative formula must be used. Most of the assumptions have to do with the types of variables and measures used and those cannot be violated. You must always have two independent groups measured on the same, quantitative variable. However, it is possible to proceed with the independent samples *t*-test if the assumption of homogeneity of variances has been violated.

Homogeneity of variances should be checked and adjustments must be made when this assumption is violated. Homogeneity of variances can be checked using something known as a Levene's test. This is often done with the aid of SPSS software. If the Levene's test indicates that the variances are significantly uneven, an adjustment to the formula is needed. SPSS provides the results for the independent samples *t*-test twice: once in standard form (which is the default to use unless the assumption of homogeneity of variances is violated) and again with an adjustment to address heterogeneous variances, which should be used when the Levene's test shows a significant violation to the assumption. With this in mind, let's turn to how to conduct an independent samples *t*-test using SPSS, paying attention to when and how to check, and address violations of, the assumption of homogeneity of variances.

8.6: Using SPSS

As reviewed in Chapter 2, software such as SPSS can be used to expedite analyses once data have been properly entered into the program. Data need to be organized and entered into SPSS in ways that serve the analysis to be conducted. Thus, this section focuses on how to enter and analyze data for an independent samples *t*-test using SPSS. SPSS version 29 was used for this book; if you are using a different version, you may see some variation from what is shown here.

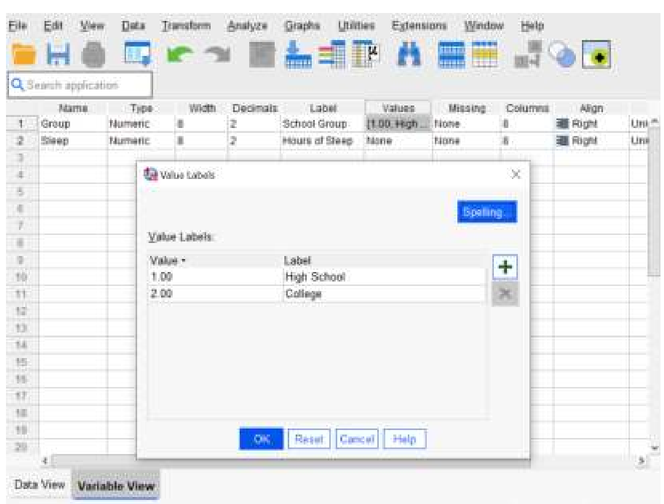
Entering Data

The independent samples *t*-tests is bivariate. One variable is used to organize data into comparison groups. The other variable is being compared between those two groups. The variable being compared must be quantitative and should have been measured using numbers on an interval or ratio scale. If these things are all true of your data, you are ready to open SPSS and begin entering your data.

Open the SPSS software, click "New Dataset," then click "Open" (or "OK" depending on which is shown in the version of the software you are using). This will create a new blank spreadsheet into which you can enter data. There are two tabs which appear towards the bottom of the spreadsheet. One is called "Variable View" which is the tab that allows you to tell the software about your variables. The other is called "Data View" which is the tab that allows you to enter your data.

Click on the Variable View tab. This tab of the spreadsheet has several columns to organize information about the variables. The first column is titled "Name." Start here and follow these steps:

1. Click the first cell of that column and enter the name of your grouping variable using no spaces, special characters, or symbols. You can name this variable "Group" for simplicity. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
2. Click the first cell of the column titled "Type" and then click the three dots that appear in the right side of the cell. Specify that the data for that variable appear as numbers by selecting "Numeric." For numeric data SPSS will automatically allow you to enter values that are up to 8 digits in length with decimals shown to the hundredths place as noted in the "Width" and "Decimal" column headers, respectively. You can edit these as needed to fit your data, though these settings will be appropriate for most variables in the behavioral sciences.
3. Click the first cell of the column titled "Label." This is where you can specify what you want the variable to be called in output, including in tables and graphs. You can use spaces or phrases here, as desired. For example, you could clarify here that the variable Group refers to "School Groups" by stating as such in the label column for this variable.
4. Click on the three dots in the first cell of the column titled "Values." This is where you can add details about each group. Click the plus sign and specify that the value 1 (for Group 1) refers to the High School subsample. Then click the plus sign again and specify that value 2 (for Group 2) refers to the College subsample as shown below. Then click "OK."



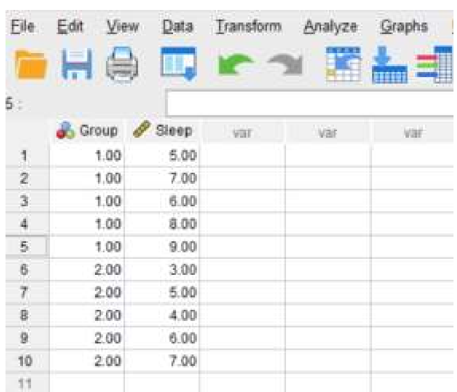
5. Click on the first cell of the column titled "Measure." A pulldown menu with three options will allow you to specify the scale of measurement for the variable. Select the "Nominal." option because grouping variables are nominal. Now SPSS is set-up for data for the grouping variable.

6. Next we need to set up space for the quantitative variable, starting with the cell under "Name," enter the name of your quantitative variable using no spaces, special characters, or symbols. You can name this variable "Sleep" for simplicity. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
7. Click the cell of the column titled "Label." You can use spaces or phrases here, as desired. For example, you could clarify here that the variable Sleep refers to "Hours of Sleep" by stating as such in the label column for this variable.
8. Click the cell of the column titled "Type" and then click the three dots that appear in the right side of the cell. Specify that the data for that variable appear as numbers by selecting "Numeric." Again, you can edit the width and decimals as needed to fit your data.
9. Click on the cell of the column titled "Measure." A pulldown menu with three options will allow you to specify the scale of measurement for the variable. SPSS does not differentiate between interval and ratio and, instead, refers to both of these as "Scale." Select the "Scale" option because if you are using an independent samples *t*-test your data for this variable should have been measured on the interval or ratio scale.

Here is what the Variable View tab would look like when created for Data Set 8.1:

Now you are ready to enter your data. Click on the Data View tab toward the bottom of the spreadsheet. This tab of the spreadsheet has several columns into which you can enter the data for each variable. Each column will show the names given to the variables that were entered previously using the Variable View tab. Click the first cell corresponding to the first row of the first column. Start here and follow these steps:

1. Enter the data for the grouping variable moving down the rows under the first column. Put a 1 for in this column for everyone who is a member of Group 1 and put a 2 for this column for everyone who is a member of Group 2.
2. Enter the data for the quantitative variable moving down the rows under the second column. If your data are already on your computer in a spreadsheet format such as excel, you can copy-paste the data in for the variable. Take special care to ensure the Sleep data for Group 1 appear in the rows corresponding to Group 1 and that the Sleep data for Group 2 appear in the rows corresponding to Group 2 as shown below. 3. Then hit save to ensure your data set will be available for you in the future.

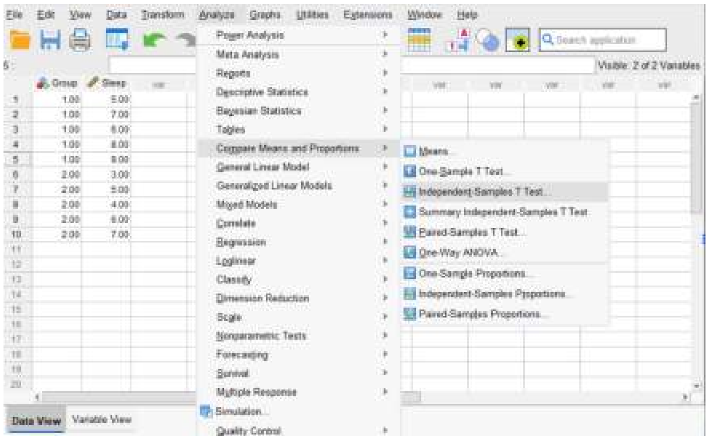


Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

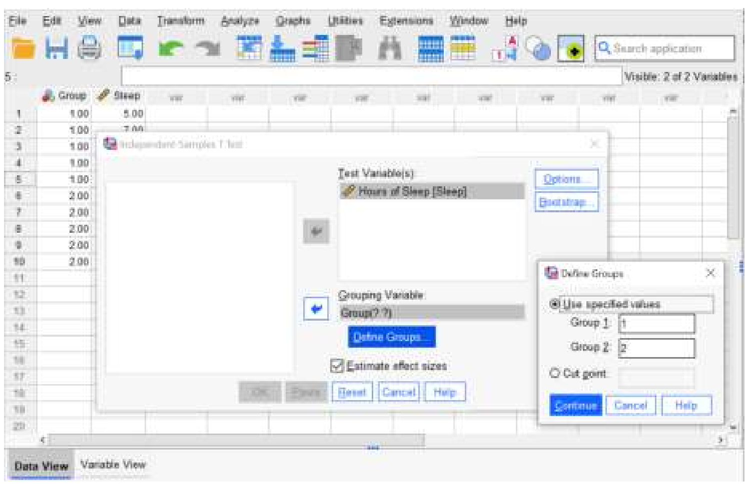
Conducting an Independent Samples t-Test in SPSS

The steps to running an independent sample *t*-test in SPSS are:

1. Click Analyze › Compare Means and Proportions › Independent-Samples T Test from the pull down menus as shown below.



2. Drag the name of the quantitative variable from the list on the left into the Test Variable(s) box on the right of the command window. You can also do this by clicking on the variable name to highlight it and the clicking the arrow to move the variable from the left into the Variable text box on the right. Next, put the grouping variable into the box bearing that name on the right side of the command window. You will need to specify what each group is named so SPSS can group the data accordingly. To do so, click the box that says “Define Groups” under the Grouping Variable section. Indicate that the Groups 1 and 2 are indicated with the numbers “1” and “2”, respectively, as shown below. Then click “Continue” on the Define Groups pop-out window. If the version of SPSS you are using has a check box to estimate effect sizes (as shown in the picture below), click that as well.



3. Click OK.
4. The output (which means the page of calculated results) will appear in a new window of SPSS known as an output viewer. The results will appear in three tables as shown below.

Group Statistics

	School Group	N	Mean	Std. Deviation	Std. Error Mean
Hours of Sleep	High School	5	7.0000	1.58114	.70711
	College	5	5.0000	1.58114	.70711

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Significance				95% Confidence Interval of the Difference	
						One Sided p	Two Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Hours of Sleep	Equal variances assumed	.000	1.000	2.000	8	.040	.081	2.00000	1.00000	-.30600	4.30600
	Equal variances not assumed			2.000	8.000.	.040	.081	2.00000	1.00000	-.30600	4.30600

Independent Samples Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	**Point Estimate**	**Lower**	**Upper**
Hours of Sleep	Cohen's d	1.58114	1.2649	-.147	2.615
	Hedges' correction	1.75156	1.1418	-.133	2.361
	Glass's delta	1.58114	1.2649	-.285	2.716

* "a": The denominator used in estimating the effect sizes. Cohen's d uses the pooled standard deviation. Hedges' correction uses the pooled standard deviation, plus a correction factor. Glass's delta uses the sample standard deviation of the control group.

Reading SPSS Output for an Independent Samples t-Test

The first table shows the descriptive statistics for the test. These include the sample size, the sample mean, the sample standard deviation, and the standard error for each group. These are versions of the six foundational pieces that would appear in the formula steps as we saw when we performed hand-calculations for Data Set 8.1 earlier in this chapter. The difference is that the table shows standard deviation whereas the formula uses variance. This is because SPSS is focusing on what you would need to interpret and report results.

The second table shows both one of the assumptions checks and the main test results which are needed for the evidence string, including the *t*-value, the degrees of freedom (*df*) and the *p* value for a one-tailed test (called "One-Sided p" in SPSS) and for a two-tailed test (called a Two Sided p in SPSS).

Let's start with the assumption check in the second table. The assumptions that must be met to use the standard form of the independent samples *t*-test is that the variances of the two groups are similar. The Levene's test for equality of variances is a homogeneity test. When variances are homogeneous enough to meet this assumption, the Levene's test will have a non-significant *p*-value (meaning that the two group variance are not significantly uneven). In the output we see the Levene's test has an obtained *F*-value of .000 and a "Sig." value (which is what SPSS calls the *p*-value) of 1.00. This *p*-value is greater than .05 indicating that the two group variances are *not* significantly uneven. This is desirable as it indicates that we have met the assumption of homogeneity of variances and can proceed to reading our results for the *t*-test.

The second output table shows two versions of the *t*-test results; those on the top row are for when the assumption of homogeneity is met and those on the bottom are for when this assumption is not met and an adjustment was employed. Our data met this assumption, so we shall use the results shown in the top row. Note: You may notice that the results for the two rows are the same for Data Set 8.1. This is because there was no impact due to unequal variances and, thus, the adjustment was not needed so it had no discernable impact.

We see that the *t*-value was 2.00, the *df* was 8, and the *p*-value was 0.04 for the one-tailed test which means the result was significant. Remember, when using the standard alpha level of .05, a *p*-value that is less than .05 is significant. These values and conclusion that the result was significant are consistent with what we found when using hand-calculations and comparing the obtained *t*-value to the critical value that fit our hypothesis and data. Therefore, the results and conclusions when using hand-calculations and when reading the results of SPSS agree. This is what will always happen unless a mistake was made in either the use of hand-calculations or in using SPSS. Note: You will sometimes see slight variation in results due to rounding error when comparing hand-calculated results to SPSS generated results. However, these differences should usually only appear in the third (thousandths) or fourth (ten-thousandths) decimal place; your hand-calculated results will usually match the SPSS generated results to the hundredths place (meaning they should match to two decimal places).

The last table of results in the SPSS output shows the effect sizes. This will only appear if you checked the box in the command window to select this extra analysis. By default, SPSS will provide three calculations of effect size: Cohen's *d*, Hedge's correction, and Glass's delta. When assumptions of the independent samples *t*-test are met and total samples sizes are sufficient, Cohen's *d* can be used. However, when total sample size is lower than 10, Hedge's *g* is preferable. The effect size is reported in the SPSS output table in the column labelled "point estimate." SPSS reports Cohen's *d* in this column as 1.2649, or 1.26 when rounded to the hundredths place for the purposes of reporting in APA format. If you compare the result for Cohen's *d* shown in the SPSS output table to our hand-calculations for earlier, you will see they are the same. You may notice that the two estimates of effect size (Cohen's vs. Hedge's) differ due to the correction used in Hedge's but are both considered large.

Choosing and Effect Size

SPSS provides three estimates of effect size, each of which is a better fit for certain scenarios. The default is to use Cohen's *d* unless small total sample size or a violation to the assumption of homogeneity of variances warrants a correction. In these cases, use of Hedge's *g* or Glass's delta are recommended, respectively.

- Cohen's *d* is appropriate when all assumptions are met and sample sizes are of sufficient size (such as when $N = 20$ or more).
- Hedge's *g* is a corrected version of Cohen's *d* which better accounts for small sample sizes. Generally, when total sample size is less than 10, Hedge's *g* is more appropriate than Cohen's *d*. When the sample size is between 10 and 20, it may be advisable to check both to see how similar or dissimilar they are.
- Glass's delta is appropriate when group variances are not homogeneous.

Reading Review 8.4

1. What scale of measurement should be indicated in SPSS for the grouping variable?
2. What scale of measurement should be indicated in SPSS for the test variable?
3. What information is used in the output to check that the assumption of homogeneity of variances is met?
4. Under which table and column of the SPSS output can the *t*-value be found?
5. Under which table and column of the SPSS output can the *p*-value be found?
6. Under which table and column of the SPSS output can the *d*-value be found?

8.7: Structured Summary for the Independent Samples t-Test

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes.

Summarize what each symbol stands for.

$\bar{X}_1$ =

$\bar{X}_2$ =

S_1^2 =

S_2^2 =

n_1 =

n_2 =

Fill-in the appropriate information for each section below:

1. Independent Samples *t*-Test Basics
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What assumptions must the data meet to use this test?
2. Independent Samples *t*-Test Formula
   a. What is the formula for an independent samples *t*-test?
   b. What are the preparatory steps for using this formula?
   c. What are the steps for solving using this formula?
3. Reporting Results from an Independent Samples *t*-Test
   a. How is this statistic reported when using APA format?
      i. What four things must be reported in the APA summary paragraph?
      ii. What two specific statistics must be reported for each of the two groups?
      iii. What are the parts of the evidence string and what does each stand for or indicate?
      iv. How is effect size computed and reported for an independent samples *t* test under standard conditions?

8.8: References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum

Navarro, D. J. (2014). Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.4). Self-published online.

CHAPTER OVERVIEW

9: Dependent Samples t-Tests

The **dependent samples *t*-test** is used when you want to test whether two means from one group measured at two different times are significantly different. This is sometimes referred to as a paired samples *t*-test or a repeated measures *t*-test. Repeated measures tests include those where the same thing is measured multiple times in the same sample. The dependent samples *t*-test is a bivariate, repeated measures test used when there are two sets of data from the same group to compare. Thus, you should use this technique if you want to compare the scores of one group to itself on a single quantitative variable. Other techniques are needed when there are two separate group being compared (such as the independent samples *t*-test reviewed in Chapter 8) or a measure is repeated three or more times (such as the repeated measures ANOVA which will be reviewed in Chapter 11).

9.1: Variables, Data, and Hypotheses that Fit the Dependent Samples t-Test

Variables

The dependent samples *t*-test is a bivariate test. This means two variables are measured and used. One of those variables must be a qualitative grouping variable used to distinguish when measurements took place (either before or after) and the other must be the quantitative variable that was measured at each of those two time points. The qualitative grouping variable is often divided into "pretest" (meaning measurements with the group were taken before something occurs) and "posttest" (meaning measurements with the group were taken after something has occurred). An example would be measuring level of sleepiness (the quantitative variable) twice and distinguishing which scores came from before coffee was consumed verses after coffee was consumed. In this example, the qualitative, grouping variable is the timing of data collection for each time sleepiness was measured. The grouping variable is used to distinguish the two sets of data for the quantitative outcome which are going to be compared.

Data

There are a few assumptions about the data which must be met before a dependent samples *t*-test is used. First, the data for the same quantitative variable must have been measured two times on the same interval or ratio scale. Second, the scores from the two waves of measurement must be matched. This means that the two scores from each time point are organized by participant so that they can appear side by side. You need to be able to identify the two scores for each person so you can see how much that participant's score changed from one wave of testing to the other. Third, there should be homogeneity of variances (though it can be okay to proceed if this assumption is not met). Finally, data for the quantitative variable should be fairly normally distributed in each group without notable impact due to outliers (such as problematic skew). However, it is worth noting that the dependent samples *t*-test is fairly robust meaning it can function well even when data are not perfectly normal and there is no need to check for homogeneity of variances (which we must check when using an independent samples *t*-test). Thus, as long as the first two assumptions are met and data are at least somewhat to fairly normal in their distributions, the dependent samples *t*-test can generally be used.

Hypotheses

Hypotheses for the dependent samples *t*-test must include both the qualitative variable and the quantitative variable and can be either non-directional or directional. Recall that directional hypotheses require one-tailed tests of significance and non-directional hypotheses should use two-tailed tests of significance. For the dependent samples *t*-test, the non-directional research hypothesis is that the scores for the sample will be different at posttest compared to at pretest. The corresponding null hypothesis is that the scores for the sample will not be different (i.e. they will be equal) at posttest compared to at pretest. Because this research hypothesis is non-directional, it requires a two-tailed test. The non-directional research and corresponding null hypotheses can be summarized as follows:

Non-Directional Hypothesis for a Dependent Samples t-Test

Research hypothesis	The posttest mean will not be equal to the pretest mean.	$H_A : \mu_{\text{post}} \neq \mu_{\text{pre}}$
Null hypothesis	The posttest mean will be equal to the pretest mean.	$H_0 : \mu_{\text{post}} = \mu_{pre}$

There are two directional hypotheses possible for the dependent samples *t*-test. One possible directional research hypothesis is that the posttest mean will be *greater than* the pretest mean. The corresponding null hypothesis is that the posttest mean will *not* be greater than the pretest mean. This could mean that the posttest mean is less than or that it is equal to the pretest mean. Because this research hypothesis is directional, it requires a one-tailed test. This version of the research and corresponding null hypotheses can be summarized as follows:

Directional Hypothesis for a Dependent Samples t-Test, Version 1

Research hypothesis	The posttest mean will be greater than the pretest mean.	$H_A : \mu_{\text{post}} > \mu_{\text{pre}}$
Null hypothesis	The posttest mean will not be greater than the pretest mean.	$H_0 : \mu_{\text{post}} \leq \mu_{pre}$

For the dependent samples *t*-test, the other possible directional research hypothesis is that the posttest mean will be *less than* the pretest mean. The corresponding null hypothesis is that the posttest mean will *not* be less than the pretest mean. This could mean that the posttest mean is greater than or that it is equal to the pretest mean. Because this research hypothesis is directional, it requires a one-tailed test. This version of the research and corresponding null hypotheses can be summarized as follows:

Directional Hypothesis for a Dependent Samples t-Test, Version 2

Research hypothesis	The posttest mean will be less than the pretest mean.	$H_A: \mu_{\text{post}} < \mu_{\text{pre}}$
Null hypothesis	The posttest mean will not be less than the pretest mean.	$H_0: \mu_{\text{post}} \geq \mu_{pre}$

These three version of the hypothesis are the broad form and would be refined to include the specific variable that is being compared at the two time points and some indication of what is happening between the time points. The grouping variable (which is often an independent variable) is focused on what is happening between the two time points. For example, if a researcher expected that people would have different mean levels of sleepiness after having coffee compared to before, the research and null hypotheses would be written as follows:

Non-Directional Hypothesis for a Dependent Samples t-Test

Research hypothesis	People will have different mean levels of sleepiness after having coffee compared to before.	$H_A: \mu_{\text{post}} \neq \mu_{\text{pre}}$
Null hypothesis	People will not have different mean levels of sleepiness after having coffee compared to before.	$H_0: \mu_{\text{post}} = \mu_{pre}$

In this example, coffee is being consumed between the two times sleepiness is measured. Thus, the quantitative comparison (or test) variable is sleepiness and the qualitative grouping variable is coffee consumption. Note that the data are grouped with regards to coffee consumption as data before (pretest) and data after (posttest).

It is also possible to propose a directional hypothesis for this scenario. For example, if the researcher expected that people would be less sleepy after having coffee compared to before they had coffee, the research and null hypotheses would be directional and could be written as follows:

Directional Hypothesis for a Dependent Samples t-Test

Research hypothesis	People will have lower mean levels of sleepiness after having coffee compared to before.	$H_A: \mu_{\text{post}} < \mu_{\text{pre}}$
Null hypothesis	People will not have lower mean levels of sleepiness after having coffee compared to before.	$H_0: \mu_{\text{post}} \geq \mu_{pre}$

Experimental Design and Cause-Effect

The dependent samples *t*-test, like the independent samples *t*-test, is sometimes used to analyze data from an experiment. When an experimental design is used and other features are present (such as temporal precedence), it may be appropriate to use causal language when interpreting and/or reporting results. However, causal language should be avoided when the design is not sufficient to support a claim of cause-effect. Therefore, it is best to use non-causal language as a default and to only switch to using causal language when it is known that an experimental design was used and that causal language is appropriate (see Chapter 8 for a brief review of experimental designs and causal language).

Reading Review 9.1

1. Is the dependent samples *t*-test univariate, bivariate, or trivariate?
2. What does the grouping variable distinguish between in a dependent samples *t*-test?

3. What is the general research hypothesis that can be tested using a one-tailed, dependent samples *t*-test when posttest scores are expected to be lower than pretest scores? Provide both sentence and symbol formats.

9.2: The Dependent Samples t-Test Formula

Before we can use the formula, it is important to understand what it can tell us and how it gets there. The dependent samples *t*-test formula looks fairly simple compared to some of the other inferential formulas we see in statistics. The obtained *t*-value tells us how large the difference is between pretest and posttest scores using standard error. Another way to say this is that it tells us how many standard errors apart the scores were at the two times of data collection. It does this by taking the difference in the scores from pretest to posttest and dividing that by the standard error of those differences. The difference is calculated in the numerator of the formula and the standard error is calculated in the denominator of the formula. Therefore, we can understand the formula's main construction and outcomes as follows:

$$t = \frac{\text{difference from pretest to postest}}{\text{standard error of the difference}}$$
$$= \text{how many standard errors of difference are observed from pretest to postest}$$

The formula contains two symbols (n and d) and the six basic mathematical operations (adding, subtracting, squaring, square rooting, multiplying, and dividing). The two symbols n and d represent, and thus must be replaced with, specific values. The dependent samples *t*-test formula is as follows:

$$t = \frac{\Sigma d}{\sqrt{\left[\frac{n\left(\Sigma d^2\right) - (\Sigma d)^2}{n-1}\right]}}$$

The two symbols are n which stands for sample size, and d which stands for difference from pretest to posttest. Difference (d) is calculated by subtracting each pretest score from its corresponding posttest score to see how different those two values are for each case. Remember, each case refers to data from one participant.

Notice that the dependent samples *t*-test formula is not asking for means. Instead, the focus is on difference from oneself. The numerator asks for the sum of differences. The denominator asks for the sample size and some additional calculations using difference scores. Thus, the formula requires we know three basic things: sample size (there is only one group so there is only one sample size), the sum of differences, and the sum of the squared differences. We also see the df in the bottom of the formula; the df for a dependent samples *t*-test is just $n-1$ because there is only one sample of participants being used to produce the two groups of data (i.e. data from that sample at pretest and data from that same sample at posttest).

Interpreting Obtained t-Values

Obtained *t*-values have two components: a magnitude and a direction. The magnitude is the absolute value of *t* in this test; this value represents how many standard errors the posttest scores were from the pretests scores, on average. Thus, the *t*-value is used to assess the difference between the posttest mean and the pretest mean, though we don't overtly see the means in the formula. The larger the *t*-value, the farther apart the pretest and posttest scores were. As the *t*-value increases, the evidence for the research hypothesis and against the null hypothesis also increases. Conversely, as the *t*-value decreases, the evidence for the research hypothesis and against the null hypothesis also decreases. Thus, researchers are generally hoping for larger *t*-values.

The other component of *t* is its direction. When *t* is positive, it indicates that posttest scores were higher than pretest scores, on average. Conversely, when *t* is negative, it indicates that posttest scores were lower than pretest scores, on average. Remember, when testing a two tailed (non-directional) hypothesis, only the magnitude needs to be considered to determine whether a result is statistically significant. However, when testing a one-tailed (directional) hypothesis, both magnitude and direction need to be considered to determine whether a result is significant. The direction of the results must match the direction of the hypothesis when using a one-tailed test of significance. For example, if it was hypothesized that posttest scores would be higher than pretests scores, the hypothesis would only be significantly supported if the differences were sufficiently large (meaning *t* exceeded the critical value) and the result was positive. Conversely, if it was hypothesized that posttest scores would be lower than pretests scores, the hypothesis would only be significantly supported if the differences were sufficiently large and the result was negative.

Note

This interpretation of direction only applies when d is computed as posttest score minus pretest score. The formula, however, works well and will produce the same magnitude of *t* value if d is computed as pretest score minus posttest score. When this occurs, a positive result means pretest scores tended to be higher and a negative result means posttest scores tended to be higher.

Reading Review 9.2

1. What is being calculated and represented by the numerator of the dependent samples *t* test formula?
2. What two things must be checked when determining whether a result from a one-tailed dependent samples *t*-test was statistically significant?

Formula Components

Now that we have taken some time to understand the construction of the dependent samples *t*-test formula, let's focus on how to actually use it, starting with identifying its parts.

In order to solve for *t*, three things must first be known:

n = the sample size

Σd = the sum of differences

Σd^2 = the sum of squared differences

To find the difference scores, subtract the pretest score from the posttest score for each case. This can be summarized as follows:

$$d = X_{\text{post}} - X_{\text{pre}}$$

Keep in mind that the symbols in the difference formula are X (which refers to an individual raw score not $\bar{x}$(which would refer to the mean). Thus, differences are calculated separately for each case using this formula before they are squared and/or summed for use in the dependent samples *t*-test formula.

Formula Steps

The steps are shown in order and categorized into two sections: A) preparation and B) solving. I recommend using this categorization to help you organize, learn, and properly use all inferential formulas. Preparation steps refer to any calculations that need to be done before values can be plugged into the formula. For the dependent samples *t*-test this includes finding the three components of the formula: n, Σd, and Σd^2. Once those are known, the steps in section B can be used to yield the obtained value for the formula. The symbol for the obtained value for each *t*-test is *t*. Follow these steps, in the specified order, to find *t*.

Section A: Preparation

1. Find n for the sample.
2. Find Σd:
   a. Find n for each member of the sample by subtracting their pretest score from their posttest score.
   b. Then, sum all the difference scores.
3. Find Σd^2 by squaring each difference score and then summing those squared values.

Section B: Solving

1. Write the formula with the values found in section A plugged into their respective locations. The numerator is completed as part of this step so we can move on to the denominator.
2. Solve for the denominator as follows:
   a. Multiply the sample size by the sum of squared differences as shown in the upper left section of the denominator.
   b. Square the sum of differences as shown in the upper right section of the denominator.
   c. Subtract the squared sum of differences (the result of Step 2b) from the sum of squared differences which has been weighted by the sample size (the result of Step 2a) to complete the top section of the denominator.
   d. Find the df by subtracting 1 from the sample size, as shown in the bottom of the denominator
   e. Divide the top part of the denominator (the result of Step 2c) by the bottom of the denominator (the result of Step 2d).
   f. Square root the results of step 2e to get the standard error of the difference. This completes the steps for the denominator.
3. Divide the sum of differences (the numerator) by the standard error of the differences (the denominator which was completed in step 2f) to get the obtained *t* value.

9.3: Distinguishing Parts of the Dependent Samples t-Test Formula

Distinguishing Parts of the Dependent Samples **t-Test Formula**

Though the dependent samples *t*-test formula is one of simpler inferential formulas, it can still cause some confusion. One mistake that is sometimes made is confusing these two components:

$(\Sigma d)^2$ = the squared sum of differences

Σd^2 = the sum of squared differences

These look and sound quite similar but order of operations dictates a different order to their steps and, thus, these are not the same thing. Here is a reminder of order of operations as it applies to each of these so you can note the distinction between them:

Component	**Steps to Solve**
$(\Sigma d)^2$ = the squared sum of differences	1. Find each difference by subtracting each pretest score from its posttest score. 2. Sum the differences. 3. Square the sum of differences.
Σd^2 = the sum of squared differences	1. Find each difference by subtracting each pretest score from its posttest score. 2. Square each difference. 3. Sum the squared differences.

Example of How to Test a Hypothesis by Computing t

Let us assume that a researcher believed that employees would have different levels of motivation after a pizza party compared to before. Suppose that motivation was measured on a scale of 1 to 10 where higher scores indicate greater motivation in a sample of 10 employees and that data were collected twice: once the day before the company sponsored pizza party (which is the pretest) and again the day after the party (which is the posttest). Assume that Data Set 9.1 includes data from the two waves of measurement. Keep in mind that each participant (i.e. case) will have two scores when using a dependent samples *t*-test: a pretest score and a posttest score. Each row contains the data from a single case. Let's use this information to follow the steps in hypothesis testing.

Data Set 9.1. Motivation Scores (n = 10)

Posttest	Pretest
6.00	7.00
5.00	4.00
4.00	5.00
7.00	6.00
5.00	5.00
4.00	4.00
6.00	5.00
3.00	6.00
5.00	2.00
8.00	7.00

Note

Putting posttest scores in the first column and pretest scores in the second can make calculations of *d* easier. However, it is also fine to show pretest scores first to reflect the timing of the waves of data collection.

Steps in Hypothesis Testing

In order to test a hypothesis, we must follow these steps:

1. **State the hypothesis.**

A non-directional hypothesis is the best fit because the goal is to see if there is a *difference* in motivation without specifying whether it will be higher or lower at posttest. A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that the mean motivation will be different after a pizza party compared to before.* If the format shown in the table below is used instead, it must be made clear that what is being stated is a research hypothesis and not a result.

Non-Directional Hypothesis for a Dependent Samples t-Test

Research hypothesis	The mean motivation after a pizza party will not be equal to the mean motivation before the pizza party.	$H_A : \mu_{\text{post}} \neq \mu_{\text{pre}}$
Null hypothesis	The mean motivation after a pizza party will be equal to the mean motivation before the pizza party.	$H_0 : \mu_{\text{post}} = \mu_{\text{pre}}$

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The scores of two dependent groups of data from the same sample are being compared so the appropriate test is a dependent samples *t*-test.

3. **Determine the critical value.**

In order to determine the critical value, three things must be identified: 1. the alpha level, 2. whether the hypothesis requires a one-tailed test or a two-tailed test, and, 3. the degrees of freedom (df).

The alpha level is often set at .05 unless there is reason to adjust it such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The hypothesis is non-directional so a two-tailed test should be used.

The df must also be calculated. Each inferential test has a unique formula for calculating df. The formula for df for the dependent samples *t*-test is as follows: $n - 1$. It appears in the bottom of the dependent samples *t*-test formula. There is only one sample being tested twice so there is only one sample size to consider. We see there are 10 cases in Data Set 9.1 so $n - 10$. Thus, df = 10 – 1 so the df for this scenario is 9.

These three pieces of information are used to locate the critical value from the test. The full tables of the critical values for *t*-tests are located in Appendix D. Below is an excerpt of the section of the *t*-tables that fits the current hypothesis and data. Under the conditions of and alpha level of .05, a two-tailed test, and 9 degrees of freedom, the critical value is 2.262.

Degrees of Freedom	**two-tailed test**		
	alpha level:	**α = 0.05**	**α = 0.01**
	9	2.262	3.250

The critical value represents the threshold of evidence needed to be confident a hypothesis is likely true. The obtained value (which is called *t* in a *t*-test) is the amount of evidence present. When using a two-tailed test, only the absolute value of the critical value must be considered. Thus, in order for the result to significantly support the hypothesis in this example, the absolute value of *t* needs to exceed the critical value of 2.262.

Degrees of Freedom for a Dependent Samples **t-Test**

Degrees of freedom (df) tell how much information you have that is free to vary. The degrees of freedom for a dependent samples *t*-test is equal to the sample size minus 1. It appears in the denominator of the dependent samples *t*-test like this:

$$n-1$$

This reflects the sample size minus the number of unique groups. Because only one sample is being tested in a dependent samples *t*-test, there is only one *n* to consider and only one time the “subtract 1” adjustment must be used.

4. **Calculate the test statistic.**

A test statistic can also be referred to as an obtained value. The formula needed to find the test statistics *t* for this scenario is as follows:

$$t=\frac{\Sigma d}{\sqrt{\left[\frac{n\left(\Sigma d^{2}\right)-(\Sigma d)^{2}}{n-1}\right]}}$$

<u>Section A: Preparation</u>

Start each inferential formula by identifying and solving for the pieces that must go into the formula. For the dependent samples *t*-test, this preparatory work is as follows:

1. Find n.

This value is found using Data Set 9.1 and is summarized as n = 10

2. Find Σd for each member of the sample by subtracting their pretest score from their posttest score and then summing those values.

The d column shows the result of subtracting the pretest score from the posttest score for each row. The total of these values is shown at the bottom of the table.

Posttest	Pretest	*d*
6.00	7.00	-1.00
5.00	4.00	1.00
4.00	5.00	-1.00
7.00	6.00	1.00
5.00	5.00	0.00
4.00	4.00	0.00
6.00	5.00	1.00
3.00	6.00	-3.00
5.00	2.00	3.00
8.00	7.00	1.00
		Σd = 2.00

This value is found using Data Set 9.1 and is summarized as Σd = 2.00

3. Find Σd^2 by squaring each difference score and then summing those values.

The d^2 column shows each d value after it has been squared. Keep in mind that negative numbers become positive when they are squared. The total of these squared d values is shown at the bottom of the table.

Posttest	Pretest	d	d^2
6.00	7.00	-1.00	1.00
5.00	4.00	1.00	1.00
4.00	5.00	-1.00	1.00
7.00	6.00	1.00	1.00
5.00	5.00	0.00	0.00
4.00	4.00	0.00	0.00
6.00	5.00	1.00	1.00
3.00	6.00	-3.00	9.00
5.00	2.00	3.00	9.00
8.00	7.00	1.00	1.00
			Σd^2 = 24.00

This value is found using Data Set 9.1 and is summarized as Σd^2 = 24.00

Now that the pieces needed for the formula have been found, we can move to Section B.

<u>Section B: Solving</u>

Now that the preparatory work is done, the formula can be used to compute the obtained value. For the dependent samples *t*-test, this work is as follows:

1. Write the formula with the values found in section A plugged into their respective locations.

Writing the formula first in symbol format before filling it in with the values can help you recognize and memorize it. Here is the formula with the symbols:

$$t=\frac{\Sigma d}{\sqrt{\left[\frac{n\left(\Sigma d^{2}\right)-(\Sigma d)^{2}}{n-1}\right]}}$$

Here is the formula with values filled into their appropriate locations in place of their symbols:

$$t=\frac{2.00}{\sqrt{\left[\frac{10(24.00)-(2.00)^{2}}{10-1}\right]}}$$

2. Solve for the denominator as follows:

> **Note**
>
> Steps will appear in bold to show when they have occurred.

a. Multiply the sample size by the sum of squared differences as shown in the upper left section of the denominator.

$$t=\frac{2.00}{\sqrt{\left[\frac{\mathbf{240.00}-(2.00)^{2}}{10-1}\right]}}$$

b. Square the sum of differences as shown in the upper right section of the denominator.

$$t=\frac{2.00}{\sqrt{\left[\frac{240.00-\mathbf{4.00}}{10-1}\right]}}$$

c. Subtract the squared sum of differences (the result of Step 2b) from the squared sum of deviations which has been weighted by the sample size (the result of Step 2a) to complete the top section of the denominator.

$$t=\frac{2.00}{\sqrt{\left[\frac{\mathbf{236.00}}{10-1}\right]}}$$

d. Find the df by subtracting 1 from the sample size, as shown in the bottom of the denominator.

$$t=\frac{2.00}{\sqrt{\left[\frac{236.00}{\mathbf{9}}\right]}}$$

e. Divide the top part of the denominator (the result of Step 2c) by the bottom of the denominator (the result of Step 2d).

$$t=\frac{2.00}{\sqrt{[\mathbf{26.2222\ldots}]}}$$

f. Square root the results of step 2e to get the standard error of the difference. This completes the steps for the denominator.

$$t=\frac{2.00}{\mathbf{5.1207\ldots}}$$

3. Divide the sum of differences (the numerator) by the standard error of the differences (the denominator which was completed in step 2f) to get the obtained *t* value, as follows:

$$t=\frac{2.00}{5.1207\ldots}$$
$$t=0.3905\ldots$$
$$t\approx 0.39$$

The final result can be rounded to the hundredths place. This result, known as a test statistic or *t*-value, can also be referred to by the general term “obtained value.” This result is positive meaning that posttest scores were higher than pretest scores, on average. However, the magnitude of the obtained value is quite low so the differences from pretest to posttest, overall, were small.

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value for t exceeds the critical value as follows: The critical value is 2.262.

The obtained *t*-value is 0.39

The obtained *t*-value does *not* exceed (i.e. is less than) the critical value. Therefore, the result is *not* statistically significant and does not support the hypothesis.

Note

We only needed to check whether the magnitude exceeded that of the critical value to determine whether the result was significant because the hypothesis was non directional and, thus, required a two-tailed test of significance.

6. **Calculate the effect size and/or other relevant secondary analyses.**

When it is determined that the result is significant, an effect size should be computed. However, when a result is not significant, effect size is not particularly useful and is generally not reported. However, we will practice how to calculate it for a dependent samples *t*-test here, despite the fact that the result was not significant, so we can learn how to compute it.

The effect size that is appropriate for t-tests under desirable conditions is known as Cohen's d (Cohen, 1988). The version of the formula varies depending upon the test used. The formula for Cohen's d is as follows when working with a dependent samples t test:

$$d = \frac{\bar{X}_d}{S_d}$$

Cohen's d, when used for a dependent samples t-test, requires two parts: the mean of the differences and the standard deviation of the differences. The mean difference is divided by the standard deviation of the differences to yield the effect size.

First, we must find the mean of differences using the following formula:

$$\bar{X}_d = \frac{\Sigma d}{n}$$

The sum of differences was found earlier to be 2.00. We can see the differences (d) in the third column below and the sum of those differences at the bottom of that column.

Posttest	Pretest	d
6.00	7.00	-1.00
5.00	4.00	1.00
4.00	5.00	-1.00
7.00	6.00	1.00
5.00	5.00	0.00
4.00	4.00	0.00
6.00	5.00	1.00
3.00	6.00	-3.00
5.00	2.00	3.00
8.00	7.00	1.00
		Σd = 2.00

The sample size is 10. We must plug those into the formula to find the mean of differences as follows:

$$\bar{X}_d = \frac{\Sigma d}{n} = \frac{2.00}{10} = 0.20$$

Next, we must find the standard deviation of the differences. The formula to do so is as follows:

$$S_d = \sqrt{\frac{\sum\left(d - \bar{X}_d\right)^2}{n-1}}$$

To use this standard deviation formula we must follow these steps:

a. Find the difference from each d value and the mean of d values. This is shown in the second column below.
b. Find the squared difference between each d value and the mean of d values. This is shown in the third column below.
c. Sum the squared deviations. This is shown in the bottom of the third column.

d	$d - \bar{X}_d$	$(d - \bar{X}_d)^2$
-1.00	-1.00 – 0.20 = -1.20	(-1.20)^ = 1.44
1.00	1.00 – 0.20 = 0.80	(0.80^2) = 0.64
-1.00	-1.00 – 0.20 = -1.20	(1.20)^ = 1.44

d	$d-\bar{X}_d$	$(d-\bar{X}_d)^2$
1.00	1.00 – 0.20 = 0.80	(0.80^2) = 0.64
0.00	0.00 – 0.20 = -0.20	$(0.20)^2$ = 0.04
0.00	0.00 – 0.20 = -0.20	$(0.20)^2$ = 0.04
1.00	1.00 – 0.20 = 0.80	(0.80^2) = 0.64
-3.00	-3.00 – 0.20 = -3.20	\((-3.20) = 10.24
3.00	3.00 – 0.20 = 2.80	\((2.80) = 7.84
1.00	1.00 – 0.20 = 0.80	(0.80^2) = 0.64
		$(d-\bar{X}_d)^2$ = 23.60

Next, divide the sum of deviations by the adjusted sample size. Then, square root to find the standard deviation as follows:

$$S_d=\sqrt{\frac{\Sigma\left(d-\bar{X}_d\right)^2}{n-1}}=\sqrt{\frac{23.60}{10-1}}=\sqrt{\frac{23.60}{9}}=\sqrt{2.6222\ldots}=1.6193\ldots$$

Now we can put the pieces together to find the effect size as follows:

$$d=\frac{\bar{x}_d}{s_d}$$
$$d=\frac{0.20}{1.6193\ldots}$$
$$d=0.1235\ldots$$

Effect sizes, like most values with decimals, are often rounded and reported to the hundredths place. Thus, this effect size is reported as $d=0.12$. Cohen's *d* can be interpreted using the following rules of thumb (Cohen, 1988; Navarro, 2014):

Interpreting Cohen's d Effect Sizes

~0.80	Large effect
~0.50	Moderate effect
~0.20	Small effect

The rules of thumb are general guidance and do not dictate precise or required interpretations. However, the rules of thumb are useful in providing an initial guideline for interpreting effect sizes. Following these rules of thumb, the current finding of d = 0.12 would be considered a small effect, at best. This is unsurprising because the result was not significant so we should not expect a particularly large effect size.

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

Results should be reported in past tense.

Finally, APA format requires a specific format be used for reporting the results of a test. The descriptive statistics needed are the means and standard deviations for the posttest and for the pretest. The steps to computing these descriptive statistics are not shown in this chapter, though their values are reported in the APA formatted summary example. For a review of how to calculate means, see Chapter 3. For a review of how to calculate standard deviations, see Chapter 4. The information needed from the inferential tests includes the degrees of freedom, obtained value, and the *p*-value.

Following this, the results for our hypothesis with Data Set 9.1 can be written as shown in the summary example.

APA Formatted Summary Example

A dependent samples *t*-test was used to test the hypothesis that the mean motivation would be different after a pizza party compared to before. Contrary to the hypothesis, the mean motivation was not significantly different at posttest (M = 5.30; SD = 1.49) compared to pretest (M = 5.10; SD = 1.52), t(9) = 0.39, $p > .05$. The Cohen's d effect size of 0.12 was very small.

This succinct summary in APA format provides a lot of detail and uses specific symbols in a particular order. To understand how to read and create a summary like this, review the detailed walk-though in Chapter 7. For a brief review of the structure for APA format, see the summary below. Note: When a result is not significant, it is not necessary to report effect sizes. However, it is included in the summary example to show how it would be reported and interpreted for this inferential test, if desired.

Summary of APA-Formatted Results for the Dependent Samples t-Test

In your APA write-up for a dependent samples *t*-test you should state:

1. Which test was used and the hypothesis which warranted its use.
2. Whether the aforementioned hypothesis was supported or not. To do so properly, four components must be reported:
   a. The mean and standard deviation for both the posttest scores and the pretest scores
   b. The test results in an evidence string as follows: *t*(*df*) = obtained value
   c. The significance portion of the evidence string as $p > .05$ if significant or $p > .05$ if not significant
   d. The effect size, if the result was significant

Anatomy of an Evidence String

The following breaks down what each part represents in the evidence string for Data Set 9.1:

Symbol for the test	Degrees of Freedom	Obtained Value	*p*-Value
t	(9)	= 0.39,	$p > .05$

Note

When a result is not significant, the significance portion of the evidence string can be written as "$p > .05$, *ns*." The *ns* portion is shorthand for "not significant" which can be helpful for audiences who are less familiar with APA-formatted summaries.

Reading Review 9.3

1. What does d stand for when computed as part of a dependent samples *t*-test?
2. What is the difference between the steps for calculating$(\Sigma d)^2$ and Σd^2?
3. Which descriptive statistics are calculated for the APA-formatted summary which are not used in the dependent samples *t*-test formula?
4. What set of symbols is used to indicate an inferential result was not statistically significant in an APA-formatted summary?
5. Under what conditions is reporting Cohen's d unnecessary?

9.4: Using SPSS

As reviewed in Chapter 2, software such as SPSS can be used to expedite analyses once data have been properly entered into the program. Data need to be organized and entered into SPSS in ways that serve the analysis to be conducted. Thus, this section focuses on how to enter and analyze data for a dependent samples *t*-test using SPSS. SPSS version 29 was used for this book; if you are using a different version, you may see some variation from what is shown here.

Entering Data

The dependent samples *t*-tests is bivariate but the way data are entered is different than for an independent samples *t*-test. Instead of entering a grouping variable, posttest scores are entered as one variable and pretest scores are entered as a second variable in SPSS. The variable being measured twice (one at posttest and once a pretest) and compared must be quantitative and should have been measured using numbers on an interval or ratio scale. If this is true of your data, you are ready to open SPSS and begin entering your data.

Open the SPSS software, click “New Dataset,” then click “Open” (or “OK” depending on which is shown in the version of the software you are using). This will create a new blank spreadsheet into which you can enter data. There are two tabs which appear towards the bottom of the spreadsheet. One is called “Variable View” which is the tab that allows you to tell the software about your variables. The other is called “Data View” which is the tab that allows you to enter your data.

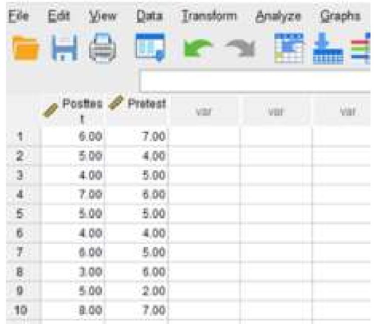
Click on the Variable View tab. This tab of the spreadsheet has several columns to organize information about the variables. The first column is titled “Name.” Start here and follow these steps:

1. Click the first cell of that column and enter the name for your posttest version of the variable using no spaces, special characters, or symbols. You can name this variable “PosttestMotivation” or just “Posttest” for simplicity. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
2. Click the first cell of the column titled “Type” and then click the three dots that appear in the right side of the cell. Specify that the data for that variable appear as numbers by selecting “Numeric.” For numeric data SPSS will automatically allow you to enter values that are up to 8 digits in length with decimals shown to the hundredths place as noted in the “Width” and “Decimal” column headers, respectively. You can edit these as needed to fit your data, though these settings will be appropriate for most quantitative variables in the behavioral sciences.
3. Click the first cell of the column titled “Label.” This is where you can specify what you want the variable to be called in output, including in tables and graphs. You can use spaces or phrases here, as desired. For example, you could clarify here that the variable Posttest refers to “Posttest Motivation” by indicating as such in the label column for this variable.
4. Click on the first cell of the column titled “Measure.” A pulldown menu with three options will allow you to specify the scale of measurement for the variable. Select the “scale.” option because the variable was measured on the interval or ratio scale. Now SPSS is set-up for data for the posttest scores of the main test variable.
5. Move to the second row of the variable table and repeat steps 1 through 4, specifying the variable as Pretest and the label as “Pretest Motivation.” Once this is done, SPSS will be set-up for data for the pretest scores of the main test variable.

Here is what the Variable View tab would look like when created for Data Set 9.1:

Now you are ready to enter your data. Click on the Data View tab toward the bottom of the spreadsheet. This tab of the spreadsheet has several columns into which you can enter the data for each variable. Each column will show the names given to the variables that were entered previously using the Variable View tab. Click the first cell corresponding to the first row of the first column. Start here and follow these steps:

1. Enter the data for the posttest version of the variable moving down the rows under the first column.
2. Enter the data for the pretest version of the variable moving down the rows under the second column. Take special care to ensure the data for each case appear next to each other. This means that each posttest score must appear next to its pretest score in the same order in which they appear in the original version of Data Set 9.1 in this chapter.
3. Then hit save to ensure your data set will be available for you in the future. Here is how Data Set 9.1 looks entered into SPSS:

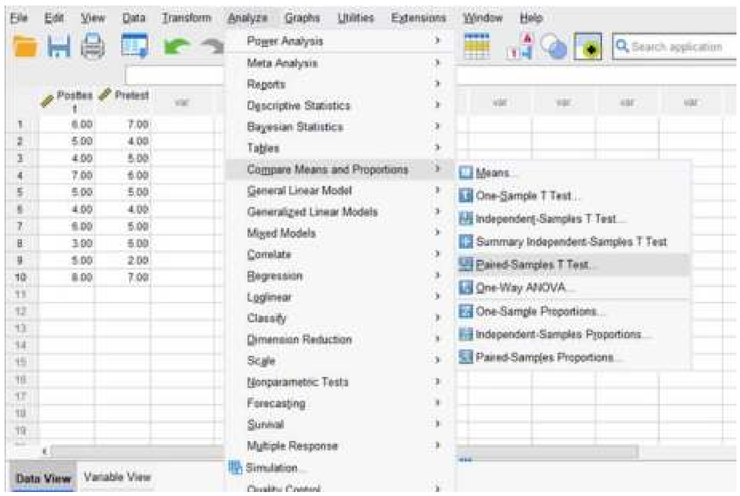


Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

Conducting a Dependent Samples t-Test in SPSS

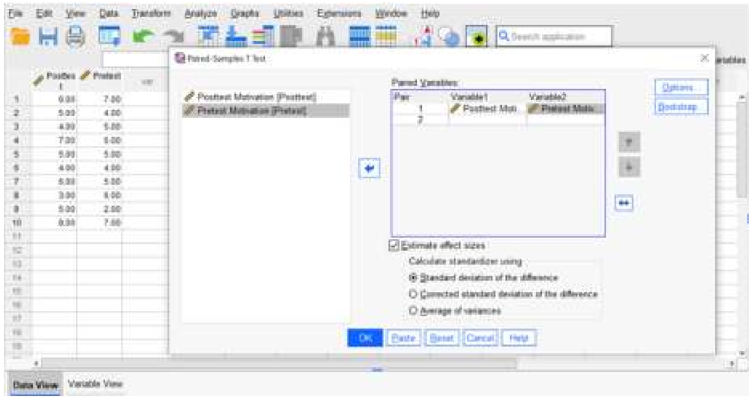
The steps to running a dependent sample *t*-test in SPSS are:

1. Click Analyze -> Compare Means and Proportions -> Paired-Samples T Test from the pull down menus as shown below.



2. Drag the name of the posttest version of the variable from the list on the left into the section for Variable 1 in the Paired Variables box on the right of the command window. You can also do this by clicking on the variable name to highlight it and the clicking the arrow to move the variable from the left into the Variable 1 text box on the right. Next, drag (or use the arrow to move) the name of the pretest version of the variable from the list on the left into the section for Variable 2 in the Paired Variables box on the right of the command window. If the version of SPSS you are using has a check box to estimate effect sizes, click that as well and select the "Standard deviation of the difference" option.

> **Note**
>
> If you reverse their order and put the pretest data as Variable 1 and the posttest data as Variable 2, it is not problematic. It will simply reverse their order in the formula causing the sign of the result to switch from either positive to negative or negative to positive. However, you could still look at the means for posttest and pretest to determine which, if either, was higher; thus, the sign of the *t*-value is not a crucial element in determining whether a directional hypothesis was supported using a dependent samples *t*-test.



3. Click OK.
4. The output (which means the page of calculated results) will appear in a new window of SPSS known as an output viewer. The results will appear in four tables as follows:

Paired Samples Statistics: Pair 1

	Mean	N	Std. Deviation	Std. Error Mean
Posttest Motivation	5.3000	10	1.49443	.47258
Pretest Motivation	5.1000	10	1.52388	.48189

Paired Samples Correlations: Pair 1

	N	Correlation	One-Sided p	Two-Sided p
Posttest Motivation & Pretest Motivation	10	.424	.111	.221

Paired Samples Test: Pair 1

Paired Differences						Significance			
				95% Confidence Interval of the Difference					
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	*df*	One Sided *p*	Two Sided *p*
Posttest Motivation -Pretest Motivation	.20000	1.61933	.51208	-.95840	1.35840	.391	9	.353	.705

Paired Samples Effect Sizes: Pair 1

				95% Confidence Interval of the Difference	
		Standardizer[a]	Point Estimate	Lower	Upper
Posttest Motivation	Cohen's *d*	1.61933	.124	-.502	.742

Pretest Motivation	Hedges' correction	1.77194	.113	-.459	.679

Reading SPSS Output for a dependent samples t-Test

The first table shows the descriptive statistics for the test. These include the sample size, the sample means, the sample standard deviations, and the standard errors for the posttest scores and the pretest scores for Data Set 9.1. Note that the overall sample size is 10 because the same 10 participants who completed the pretest also completed the posttest. We will need the means and standard deviations for each wave of testing (posttest and pretest) when writing an APA-formatted results summary.

The second table shows correlation between the posttest and pretest scores. We do not need these for the standard way we are using the dependent samples *t*-test in this chapter so we can move to the next table of results.

The third output table shows the *t*-test results. Though the table lists several things, the main three items we need are the degrees of freedom, the obtained *t*-value, and the p-value. We see that the *t*-value was 0.39 when rounded to the hundredths place, the df was 9, and the p-value was 0.705 for the two-tailed (called "two-sided" by SPSS) test which means the result was *not* significant. Remember, when using the standard alpha level of .05, a p-value that is less than .05 is significant and a p-value greater than 0.05 is not significant. These values and the conclusion that the result was not significant are consistent with what we found when using hand-calculations and comparing the obtained *t*-value to the critical value that fit our hypothesis and data. Therefore, the results and conclusions when using hand-calculations and when reading the results of SPSS agree. This is what should always happen unless a mistake was made in either the use of hand-calculations or in using SPSS. Note that you may see slight variation when comparing hand-calculations to SPSS results if you rounded your steps when doing the work by hand.

The last table of results in the SPSS output shows the effect sizes. This will only appear if you checked the box in the command window to select this extra analysis. Cohen's *d* can generally be used; however, the Hedge's correction is sometimes recommended and should be considered when working with sample sizes smaller than 20. Because our *t*-test was not significant and, thus, we are only checking the effect size for demonstration purposes, we will stick to reviewing the Cohen's d value. The effect size is reported in the SPSS output table in the column labelled "point estimate." SPSS reports Cohen's *d* in this column as 0.124, or 0.12 when rounded to the hundredths place. If you compare the result for Cohen's d shown in the SPSS output table to our hand-calculations for earlier, you will see they are the same. All is well.

Reading Review 9.4

1. What scale of measurement should be indicated in SPSS for each version of the test variable?
2. Under which table and column of the SPSS output can the *t*-value be found?
3. Under which table and column of the SPSS output can the p-value be found?
4. Under which table and column of the SPSS output can the Cohen's d-value be found?
5. Under what conditions is the Hedge's correction for effect size recommended over using Cohen's d?

9.5: Structured Summary for the Dependent Samples t-Test

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes.

Summarize what each symbol stands for.

n =

Σd =

Σd^2 =

Fill-in the appropriate information for each section below:

1. Dependent Samples *t*-Test Basics
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What assumptions must the data meet to use this test?
2. Dependent Samples *t*-Test Formula
   a. What is the formula for a dependent samples *t*-test?
   b. What are the preparatory steps for using this formula?
   c. What are the steps for solving using this formula?
3. Reporting Results from a dependent samples *t*-Test
   a. How is this statistic reported when using APA format?
      i. What four things must be reported in the APA summary paragraph?
      ii. What two specific statistics must be reported for each of the two versions of the test variable (i.e. for the posttest and pretest)?
      iii. What are the parts of the evidence string and what does each stand for or indicate?
      iv. How is the effect size computed and reported for a dependent samples *t* test with sample sizes greater than 20?
      v. What is the recommended effect size to check and report using SPSS when the sample size is smaller than 20?

9.6: References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum

Navarro, D. J. (2014). Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.4). Self-published online.

CHAPTER OVERVIEW

10: One-Way ANOVA

10.1: Introduction to One-Way ANOVA

The **one-way ANOVA** is used when you want to test whether the means of three or more independent groups are significantly different. This inferential test can also be referred to as a simple ANOVA or a single-factor ANOVA. ANOVA is short for analysis of variance because the formula takes variability into account. If the means are distinct enough even after accounting for the fact that each independent group has some variability around their own mean, the result will be significant. One-way ANOVA is a bivariate technique. One variable is a qualitative grouping variable and the other is a comparison or focal quantitative variable. Essentially, one way ANOVA is similar to an independent samples *t*-test but without the limitation of only being able to compare two groups (see Chapter 8 to review the independent samples *t*-test).

It is worth noting when and why a one-way ANOVA is used in place of an independent samples *t*-test. An ANOVA can be used to compare two independent groups but it is more complex than the *t*-test. Therefore, it is advisable to use the simpler independent samples *t*-test to compare two groups. It is also possible to use independent samples *t*-tests to compare more than two groups, but this is inefficient and introduces a greater probability of a Type I Error. Recall that a Type I Error refers to when data support a hypothesis when the hypothesis is not true of the population and that we generally take a 5% risk of this occurring each time we test a hypothesis (See Chapter 6 for a review of Type I Error). For example, three groups could be compared using three independent samples *t*-tests as follows: 1. Comparing Group 1 to Group 2, 2. Comparing Group 1 to Group 3, and 3. Comparing Group 2 to Group 3. That is a lot of work and each time the *t*-test is performed using an alpha level of 0.05, there is a 5% risk of a Type I Error. Thus, if three such tests are performed, there is a 15% chance that at least one of the three *t*-tests has a Type I Error. For these reasons, a one-way ANOVA is the preferred and appropriate method for comparing the scores of three or more independent groups.

10.2: Variables, Data, and Hypotheses that Fit a One-way ANOVA

Variables

The one-way ANOVA requires the use of one qualitative grouping variable with two or more independent groups (though ANOVA is only recommended when there are at least three independent groups) and a quantitative variable which was measured the same way in each group. An example of a qualitative grouping variable with four groups would be class standing where participants are grouped as “Freshmen,” “Sophomores,” “Juniors,” and “Seniors.” The grouping variable is used to distinguish the independent groups which are going to be compared. The thing being compared between the groups is the quantitative variable. This is sometimes considered an outcome variable. An example of a quantitative variable would be hours of sleep. In this example, all participants would have their hours of sleep measured and used in the analysis. For example, if a researcher wanted to test whether those with different class standings had different mean hours of sleep per night, the qualitative grouping variable would be class standing and the quantitative variable being compared between the four groups would be sleep hours.

Data

Each statistical test has some assumptions which must be met in order for the formula to function properly. In keeping, there are a few assumptions about the data which must be met before a one-way ANOVA is used. First, the data for the quantitative variable should have been measured the same way for all cases in each of the groups. Second, the data for the quantitative variable must be measured on the interval or ratio scales of measurement. Third the members of the groups must be non-overlapping (i.e. independent of one another). This means that no participant can be in more than one group. Fourth, data for the quantitative variable should be fairly normally distributed in each group. Finally, there should be homogeneity of variances. **Homogeneity of variances** is when the variances for the quantitative variable of each group are similar. When variances are not homogeneous it means that the groups have different amounts of error (as measured using variance) so their distribution curves have different widths and heights. When variances are not homogenous, adjustments to the formula are required. An alternative option to consider when variances are significantly different is reviewed later in this chapter. It is also helpful, though not always required, that the sample sizes of the groups be similar. When variances are similar among groups, unequal sample sizes of groups may not need to be considered. For now, if all five of these assumptions are met, the one-way ANOVA can be used.

Hypotheses

Before we can review the hypotheses which are appropriate for ANOVA, we need to review a distinguishing characteristics of ANOVA. ANOVA is often performed with a companion test. To distinguish these the ANOVA is often referred to as an omnibus test and its companion is known as a post-hoc test. The omnibus test is used to test whether there is at least one group mean that is significantly different from another. This is what the ANOVA, itself, is able to determine. However, the ANOVA cannot tell us which group(s) were different from which other group(s). Therefore, the ANOVA alone is not sufficient for testing directional hypotheses. This is why a second analysis known as post-hoc testing is sometimes needed alongside an ANOVA. Post-hoc tests are used to check each group mean against the others to determine which ones were significantly different from each other. Thus, this chapter includes both how to perform the ANOVA (omnibus test) and post-hoc tests to provide you with the skills necessary for testing directional and non-directional hypotheses. We will cover how to perform both the omnibus tests and post-hoc tests, as needed, in our computation sections later in this chapter.

Hypotheses for the one-way ANOVA must include both a qualitative grouping variable and a quantitative test variable. The non-directional hypothesis is the default as it is the only kind of hypothesis that can be tested using the ANOVA formula alone. However, directional hypotheses can also be tested provided post-hoc analyses are performed following the one-way ANOVA. For the one-way ANOVA, the non-directional research hypothesis is that the sample means will be different from each other. The corresponding null hypothesis is that the sample means will not be different from each other. Note: Because there can be more than three groups in an ANOVA, only three will be shown and ellipses will be used to indicate that the hypothesis could be expanded to include more than three groups. The non-directional research and corresponding null hypotheses can be summarized as follows:

Non-Directional Hypothesis for a One-Way ANOVA

Research hypothesis	The means of the groups are not all equal to each other.	$H_A: \mu_1 \neq \mu_2 \neq \mu_3$
Null hypothesis	The means of the groups are all equal to each other.	$H_0: \mu_1 = \mu_2 = \mu_{3\ldots}$

To test a non-directional hypothesis such as this, only the omnibus part of ANOVA is required.

There are many different directional hypotheses possible for the one-way ANOVA. We will review just one of them as an example presuming there are three groups being compared. One possible directional research hypothesis is that the mean for Group 1 will be *greater than* the mean for Group 2 and that the mean for Group 2 will be *greater than* the mean for Group 3. The corresponding null hypothesis is that the means for Group 1 through Group 3 will be less than or equal to each other, respectively. This version of the research and corresponding null hypotheses can be summarized as follows:

Example of a Directional Hypothesis for a One-Way ANOVA

Research hypothesis	The mean of Group 1 will be greater than the mean of Group 2 and the mean for Group 2 will be greater than the mean for Group 3.	$H_A : \mu_1 > \mu_2 > \mu_3$
Null hypothesis	The mean of Group 1 will not be greater than the mean of Group 2 and/or the mean for Group 2 will not be greater than the mean for Group 3.	$H_0 : \mu_1 \leq \mu_2 \leq \mu_{3\ldots}$

To test a directional hypothesis such as this, both the ANOVA and post-hoc tests are required. The omnibus can only tell us whether any of the means are different and the post-hoc adds to this by specifying which means were different and in which ways (i.e. which group means were significantly higher than others, which were significantly lower, and which were not significantly different).

It is important to note that the ANOVA omnibus formula is broad and uses a one-tailed test which is often referred to as a right-tailed test regardless of whether the hypothesis is directional or not. This is distinct from how the direction of hypotheses are used to determine whether a one-tailed or two-tailed test of the hypothesis is needed when using *t*-tests. In ANOVA, the directional aspects of a hypothesis, when present, require a secondary analysis known as post-hoc testing.

Experimental Design and Cause-Effect.

The one-way ANOVA is sometimes used to analyze data from an experiment when multiple conditions are being tested and compared. When a true experimental design is used (and other relevant conditions such as temporal precedence are met) it can be appropriate to use causal language in hypotheses and when interpreting results (see Chapter 8 for a review of experimental designs and deducing cause-effect). When ANOVA is used to test cause-effect patterns, the qualitative grouping variable is the *independent variable* and the quantitative test variable is the *dependent variable* (see Chapter 1 for a review of these two categorizations of variables). However, one-way ANOVA is flexible and can also be used to compare the means of different groups when data were not acquired using an experimental design. Thus, it is not always appropriate to use causal language with ANOVA. The grouping variable in ANOVA is, therefore, often referred to as a *factor* rather than an independent variable. It is best to use non-causal language as a default and to only switch to using causal language when it is known that an experimental design was used and that causal language is appropriate.

Reading Review 10.1

1. How many independent groups can be compared using one-way ANOVA?
2. What assumptions must be met before using a one-way ANOVA?
3. For what two reasons is one-way ANOVA preferred over using independent sample *t*-tests to compare the means of three or more independent groups?
4. When ANOVA is used to test data from an experiment, which variable generally serves as the independent variable and which generally serves as the dependent variable?
5. What is the grouping variable referred to in ANOVA when data are being tested from non experimental designs?
6. What kind of hypothesis can be fully tested using the ANOVA omnibus test?
7. What kinds of hypothesis require post-hoc testing be performed in addition to the ANOVA omnibus test?

10.3: The One-Way ANOVA Formula

There are two main processes in testing a hypothesis stating that the means of three or more groups are not all equal:

1. An ANOVA omnibus test and
2. A post-hoc test.

Omnibus means containing or representing many things at once. The first process, the omnibus test, is where the ANOVA formula is used. The omnibus test performed using the ANOVA formula is a blunt test that can tell us whether, either, all group means are approximately equal or at least one group mean is significantly different than at least one other group mean. In this part of the process, the ANOVA formula considers data from all independent groups together. When the result of the ANOVA omnibus test is significant, it indicates that not all group means are equal to each other. This test alone is sufficient to testing whether a non-directional hypothesis is or is not supported.

When an omnibus test is significant and a hypothesis is directional, additional analyses are needed to reveal which groups are significantly different from which other groups. This second round of analysis is known as post-hoc testing. In post-hoc testing, two group means are compared at a time to determine which group(s) are significantly different from which other group(s). Comparing two groups at a time is referred to as a *pairwise comparison.* However, when the result of the omnibus test is not significant, it indicates that group means are approximately equal to one another and that no further testing is warranted. Thus, post hoc tests are performed after the result of an ANOVA omnibus test is determined to be statistically significant and when a hypothesis was directional. We will tackle these in their appropriate order; therefore, we must start with the ANOVA omnibus formula and how to use it before moving on to post-hoc testing.

Components of the ANOVA Formula

The omnibus test is performed using the formula for a one-way ANOVA. Before we can use the formula, it is important to understand what it can tell us and how it gets there. The one-way ANOVA formula has two main parts. The numerator focuses on difference between groups and the denominator focuses on differences within groups. The obtained value is indicated with the symbol F . F tells us the ratio of difference between groups relative to difference within groups. Another way to say this is that it tells us how different the groups are from one another after taking into account how different members of each group are from other members of their own group. The goal is to see whether groups are more distinct from one another than individuals are from members within groups.

The variation between groups can be thought of as system differences and/or differences caused by the independent variable in an experiment plus some error. The error is the random, non-systemic and/or non-treatment related differences that occur; this random error can be estimated by calculating the variation among members within groups. Thus, the variation within groups can be thought of as noise we want to account for so we can see the how much variation is systemic and non-random. Note that the systematic variation is the estimate of how much a dependent variable was affected by an independent variable when a true experiment is used. What we want to know is how much of the variability between groups is systemic. Therefore, we can understand the formula's main construction and outcomes as follows:

$$F=\frac{\text{variation between groups}}{\text{variation within groups}}=\text{ ratio of systemic differences between groups}$$

The denominator of the formula can be referred to as the error term. In ANOVA, an **error term** is a calculation of the amount of variation that is random/non-systemic (or which is not estimated to be caused by the independent variable in a true experiment).

Unfortunately, the formula looks simpler than it is so we will look at it first then expand it out to understand its two main parts and how to calculate them. The one-way ANOVA formula is as follows:

$$F=\frac{MSS_b}{MSS_w}$$

The numerator asks for the mean sum of squares between groups (MSS_b). The denominator asks for the mean sum of squares within groups (MSS_w). Calculating each of these requires that we first calculate an SS and a df. This is because $MSS=SS\div df$ so the formula can be rewritten as follows:

$$F=\frac{SS_b\div df_b}{SS_w\div df_w}$$

The denominator of the ANOVA formula

The parts of the denominator are more familiar so we will start with those. The two components needed are SS_w and df_w.

Sum of Squares Within (SS_w)

SS_w is the sum of squared deviations within the group (also known simply as Sum of Squares Within); this has its own formula which we must use to calculate the sum of squares within each independent group. The formula for this is the same one we used in Chapter 4 in route to finding standard deviations and variances. The formula is as follows:

$$SS_w = \Sigma(x - \bar{x})^2$$

The steps to findings the SS_w for each independent group are:

1. Find the mean.
2. Subtract the mean from each raw score to find each deviation.
3. Square each deviation.
4. Sum the squared deviations

Each group has its own SS_w and we need to put them all together for ANOVA. Therefore, we must find SS_w for each group and add them together to use in the formula. If we had three groups, this could be summarized as follows:

$$SS_w = SS_{w1} + SS_{w2} + SS_{w3}$$

Degrees of Freedom Within (df_w)

df_w is the degrees of freedom within the group; this has its own formula which we must use. The formula is as follows:

$$df_w = N - k$$

The steps to findings the df_w are:

1. Find the total sample size (N) by adding the sample sizes (n) of all groups together.
2. Find k which refers to the number of independent groups or factors for the qualitative, grouping variable.
3. Subtract the number of groups (k) from the sum of sample sizes (N).

The numerator of the ANOVA formula

The two components needed for the numerator are SS_b and df_b. Each of these has its own formula which we must use. The parts of the numerator are less familiar because we have not seen them in any prior chapter of this book. Therefore, be sure to take extra care in reviewing these new formulas.

Sum of Squares Between (SS_b)

SS_b is the sum of squared deviations between groups (also known simply as Sum of Squares Between). This must be calculated for each group before they can be added together. The formula for calculating SS_b for each independent group is as follows:

$$SS_{\text{b\_group}} = n_{\text{group}} \left[(\bar{x}_{\text{group}} - \bar{x}_{\text{grand}})^2 \right]$$

We have some new pieces here so let's review all of them. The parts of the formula and their translations are as follows:

n_{group} : the sample size for a group

$\bar{x}_{\text{group}}$: the mean for a group

$\bar{x}_{\text{grand}}$: the mean when all data for all groups are treated as one grand group

The group sample sizes and means are the same ones we have used in prior chapters, however, the grand mean has not appeared in a prior chapter. The grand mean ($\bar{x}_{\text{grand}}$) is the mean for all scores together, regardless of their individual group memberships. To find the grand mean, we must sum all the raw scores for all groups to get something known as the grand total or G. Then, we divide G by the total sample size which is N. The formula for the grand mean, therefore, is as follows:

$$\bar{x}_{\text{grand}} = \frac{G}{N}$$

The steps to findings the SS_b for each independent group are:

1. Find the grand mean.
2. Subtract the grand mean from the group mean to find the deviation between these two means.
3. Square the deviation between the means.
4. Multiply the squared deviation by the size of the group.

Each group has its own SS_b and we need to put them all together for ANOVA. Therefore, we must find SS_b for each group and add them together to use in the formula. If we had three groups, this could be summarized as follows:

$$SS_b = SS_{b1} + SS_{b2} + SS_{b3}$$

The formula for sum of squared deviation between can also be written showing the steps per group and the step to sum those group values together in one formula as follows:

$$SS_b = \Sigma n_i \left[\left(\bar{x}_i - \bar{x}_{\text{grand}} \right)^2 \right]$$

The subscript i stands in for the names of all groups being tested such that the computations should each be computed for Group 1, then Group 2, then Group 3 and so on until computations for all groups have been completed. Thus, what this is saying is that, if you want to know the overall SS_b, you need to first find SS_b for each group and then sum those to get the overall SS_b.

You may have noticed some differences between the sum of squares within and the sum of squares between. Let's take a moment to consider their similarities and differences. They both find deviations and square those deviations. However, the within calculations focus on individual scores verses their group mean whereas between calculations focus on the group verses the grand group. Further, each individual in a group has their deviation overtly calculated in a within calculation (using $x - \bar{x}$); however, deviation is calculated at the group level in the between calculations (using $\bar{x}_{\text{group}} - \bar{x}_{\text{grand}}$). The group mean is representing all members of the group so the sample size needs to be taken into account to address this. This is why the squared group deviations in the SS_b formula are multiplied by their sample sizes. This adds weight to the squared deviation for each group proportional to the number of scores being represented by the group means.

Degrees of Freedom Between (df_b)

df_b is the degrees of freedom between the groups; this has its own formula which we must use. This is a new version of df. The formula is as follows:

$$df_b = k - 1$$

The steps to findings the df_b are:

1. Find k which refers to the number of independent groups or factors for the qualitative, grouping variable.
2. Subtract 1 from k.

df_b is always the number of groups minus 1. For example, if three independent groups were being compared, the df_b would be 2 but if four independent groups were being compared the df_b would be 3, and so on.

Putting the Formula Together

Once the four components are calculated, their results are put into the ANOVA formula and used to solve for F.

$$F = \frac{SS_b \div df_b}{SS_w \div df_w} = \frac{MSS_b}{MSS_w}$$

Interpreting Obtained F-Values

Obtained F-values are always positive so only their magnitude, and not their direction, is interpreted. The magnitude represents the ratio of variation that is systematic and non-random relative to the amount of variation that is non-systematic and random. Remember that the difference within groups represents the error term and is calculated in the denominator. When the F-value is 1.00 it means there is as much difference between groups as there is within groups; this indicates that all the difference observed is just random and does not represent actual differences between groups. When this occurs, the null hypothesis is retained. Thus, the closer F is to 1.00, the less the difference observed is attributed to group by group (between) variation.

Consistent with this, when the difference between the groups is greater that the error term (i.e. differences within groups), the F-value will be greater than 1.00. The larger the F-value, the more the difference observed is attributed to group by group (between) variation. Another way to say this is that the larger the F-value, the greater the non-random differences are between groups and, thus, the more evidence there is in support of the alternative hypothesis and against the null hypothesis. When the F-value is large enough to surpass the critical value, it means that the differences observed between groups (after accounting for difference within groups) is unlikely to be due to chance (i.e. it is unlikely to be random). When this occurs, the result can be declared statistically significant.

Reading Review 10.2

1. What is being calculated and represented by the numerator of the one-way ANOVA formula?
2. What is being calculated and represented by the denominator of the one-way ANOVA formula?
3. What is a grand mean and how is it calculated?

10.4: Using the ANOVA Formula

Formula Component Summary

Now that we have taken some time to understand the construction of the one-way ANOVA formula, let's focus on how to actually use it, starting with identifying all of its parts.

In order to solve for F, four things must first be known:

SS_w = sum of squares within groups

df_w = degrees of freedom within groups

SS_b = sum of squares between groups

df_b = degrees of freedom between groups

Formula Steps

The steps are shown in order and categorized into two sections:

A. preparation and
B. solving.

Preparation steps refer to any calculations that need to be done before values can be plugged into the formula. For the one-way ANOVA this includes finding several preparatory components needed for the formula (listed in the section above). Once those are known, the steps in section B can be used to yield the obtained value for the formula. The symbol for the obtained value for ANOVA is F. Follow these steps, in order, to find F.

Section A: Preparation

1. Find k (the number of independent groups being compared)
2. Find the n for each group
3. Find N (the total sample size across all independent groups being compared)
4. Find $\bar{x}$ for each group
5. Find the $\bar{x}_{\text{grand}}$ (grand mean)

Note

We will need the mean and standard deviation for each group when reporting results in APA-format. Therefore, it can be useful to compute the standard deviation for each group during the preparatory steps even though they are not needed for the ANOVA formula.

Section B: Solving

The values from the preparatory steps must now be used to find the four main components of the ANOVA formula before the F-value can be found.

1. Find the SS_w for each independent group then sum them to get the total SS_w:

$$SS_w = \Sigma(x - \bar{x})^2$$

   a. Subtract the mean for Group 1 from each raw score in Group 1 to find this groups deviations.
   b. Square the deviations for Group 1.
   c. Sum the squared deviations to get the SS_{w1} (which is the Sum of Squares Within for Group 1).
   d. Repeat steps 1a through 1c for each group until the SS_w for each group is known.
   e. Add the SS_w (Sum of Squares Within) for all groups together to get the total SS_w.
2. Find the df_w by subtracting the number of groups (k) from the total of sample sizes (N).

$$df_w = N - k$$

3. Find the total SS_b across all groups:

$$SS_b = \Sigma n_i \left[\left(\bar{x}_i - \bar{x}_{\text{grand}} \right)^2 \right]$$

a. Subtract the grand mean ($\bar{x}_{\text{grand}}$) from the mean for Group 1 ($\bar{x}_1$)
b. Square the group deviation from step 3a.
c. Multiply the squared deviation for the group (which is the result of step 3b) by the sample size for Group 1 (n_1) to find the Sum of Squares Between for Group 1 (SS_b).
d. Repeat steps 1a through 1c for each group until the SS_b for each group is known.
e. Add the SS_b (Sum of Squares Between) for all groups together to get the total SS_b.

4. Find the df_b by subtracting 1 from the number of groups (k).

$$df_b = k - 1$$

5. Write the ANOVA formula with the four values found in the above steps (i. e. SS_w, df_w, SS_b, and df_b) plugged into their respective locations.

$$F = \frac{SS_b \div df_b}{SS_w \div df_w}$$

6. Solve for MSS_b by dividing SS_b by df_b. This gives you the numerator for the F formula.

$$MSS_b = SS_b \div df_b$$

7. Solve for MSS_w by dividing SS_w by df_w. This gives you the denominator for the F formula.

$$MSS_w = SS_w \div df_w$$

8. Finally, divide MSS_b (which is the result of step 6) by MSS_w (which is the result of step 7) to get the obtained F-value.

$$F = \frac{MSS_b}{MSS_w}$$

Example of How to Test a Hypothesis by Computing F.

Let us assume that a researcher believed that children in three different conditions would differ in their mean number of aggressive behaviors toward a toy. Suppose that aggression was measured in number of physically aggressive acts (such as individual hits, kicks, and throws) towards a toy for three different groups of children. Group 1 has been shown an adult acting aggressively toward the toy (such as by hitting, kicking, and throwing it), Group 2 has been shown an adult playing non-aggressively with the toy (such as by picking it up, sitting it down, and pretending to share with it), while Group 3 is not shown the toy or any interactions with it. Assume that Data Set 10.1 includes data from the three independent samples. Let's use this information to follow the steps in hypothesis testing.

Data Set 10.1. Aggressive Acts by Children in Three Independent Viewing Groups (n = 21)

Group 1	Group 2	Group 3
5	3	0
7	1	1
6	2	0
4	4	3
8	2	1
6	2	1
6	0	1

Steps in Hypothesis Testing

In order to test a hypothesis, we must follow these steps:

1. **State the hypothesis.**

A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that the mean acts of aggression will be different among children in three different conditions.* The format shown below can also be used instead. Remember, a one-tailed test is used when conducting the omnibus test in ANOVA, regardless of whether the hypothesis is stated directionally.

Non-Directional Hypothesis for a One-Way ANOVA

Research hypothesis	The mean of the groups are not all equal to each other.	$H_A: \mu_1 \neq \mu_2 \neq \mu_3 \ldots$
Null hypothesis	The mean of the groups are all equal to each other.	$H_0: \mu_1 = \mu_2 = \mu_3 \ldots$

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The means of three independent samples are being compared so the appropriate test is a one-way ANOVA.

3. **Determine the critical value.**

In order to determine the critical value for a one-way ANOVA, three things must be identified:

1. the alpha level,
2. the Degrees of Freedom Between (df_b), and,
3. the Degrees of Freedom Within (df_w).

The alpha level is often set at .05 unless there is reason to adjust it such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The df_b and the df_w must also be calculated. These will be used to find the critical value and these will also be important pieces in the ANOVA formula. Let's find each for Data Set 10.1. There are three groups (k = 3) and each has a sample size of 7 ($n_1 = 7, n_2 = 7$, and $n_3 = 7$) The total sample size across the three groups is 21 (N = 21).

$$df_w = N - k = 21 - 3 = 18$$

$$df_b = k - 1 = 3 - 1 = 2$$

These three pieces of information are used to locate the critical value for the test. The full tables of the critical values for F-tests are located in Appendix E. Below is an excerpt of the section of the F-tables that fits the current hypothesis and data. Under the conditions of an alpha level of .05, df_b = 2, and df_w = 18, the critical value is 3.555.

Critical Values Table

		Degrees of Freedom Between (df_b)							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Degrees of Freedom Within (df_w)	**18**	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510

The critical value represents the value which must be exceeded in order to declare a result statistically significant. The obtained value (which is called F in an ANOVA) is the amount of evidence present. F-values will always be positive so we do not have to worry about any negative signs when comparing the obtained value for F to the critical value; only the magnitude of the obtained F-value must be considered. Thus, in order for the result to significantly support the hypothesis is needs to exceed the critical value of 3.555.

4. **Calculate the test statistic.**

A test statistic can also be referred to as an obtained value. The formula needed to find the test statistics, known as F for this scenario, is as follows:

$$F = \frac{MSS_b}{MSS_w}$$

But we must remember that the *MSS* in each section stands for the mean sum of squares and that each of these is actually comprised of two parts: an SS and a df. Thus, it is more helpful to break the formula out to show those pieces as follows:

$$F = \frac{SS_b \div df_b}{SS_w \div df_w} = \frac{MSS_b}{MSS_w}$$

Section A: Preparation

Start each inferential formula by identifying and solving for the pieces that must go into the formula. For the one-way ANOVA, this preparatory work is as follows:

1. Find k (the number of independent groups being compared).

This value is found using Data Set 10.1 and is summarized as k *= 3*

2. Find the n for each group.

This value is found using Data Set 10.1 for each group and is summarized as follows:

n_1 = 7

n_2 = 7

n_3 = 7

3. Find N (the total sample size across all independent groups being compared). *This value is found using Data Set 10.1 and is summarized as follows:*

$$N = n_1 + n_2 + n_3$$
$$N = 7 + 7 + 7$$
$$N = 21$$

4. Find $\bar{x}$ for each group.

These values are found using Data Set 10.1 and are summarized as follows:

$\bar{X}_1$ = 6.00

$\bar{X}_2$ = 2.00

$\bar{X}_3$ = 1.00

5. Find the $\bar{x}_{\text{grand}}$ (grand mean).

This value is found using Data Set 10.1 and is summarized as follows: 21= 63

$$\bar{x}_{\text{grand}} = \frac{5+7+6+4+8+6+6+3+1+2+4+2+2+0+0+1+0+3+1+1+1}{21} = \frac{63}{21} = 3.00$$

Now that the pieces needed for the formula have been found, we can move to Section B.

Section B: Solving

The values from the preparatory steps must now be used to find the four main components of the ANOVA formula before the F-value can be solved. The calculations for steps 1a through 1d are shown in the calculations table below.

1. Find the SS_w for each independent group then sum them to get the total SS_w:
   a. Subtract the mean for Group 1 from each raw score in Group 1 to find their deviations.
   b. Square the deviations for Group 1.
   c. Sum the squared deviations to get the SS_w1 (which is the Sum of Squares Within for group 1).
   d. Repeat steps 1a through 1c for each group until the SS_w for each group is known.

Note

The summarized calculations are shown below for reference. For a detailed review of how to calculate a sample mean, see Chapter 3. For a detailed review of how to calculate sums of squares within, see Chapter 4.

Calculations for Descriptive Statistics and SS_w **for each Group in Data Set 10.1**

Group 1	$X-\bar{X}_1$	$(X-\bar{X}_1)^2$	**Group 2**	$(X-\bar{X}_2)^2$	$(X-\bar{X}_2)^2$	**Group 3**	$X-\bar{X}_3$	$(X-\bar{X}_3)^2$
5	(-1)	1	**3**	1	1	**0**	(-1)	1
7	1	1	**1**	(-1)	1	**1**	0	0
6	0	0	**2**	0	0	**0**	(-1)	1
4	(-2)	4	**4**	2	4	**3**	2	4
8	2	4	**2**	0	0	**1**	0	0
6	0	0	**2**	0	0	**1**	0	0
6	0	0	**0**	(-2)	4	**1**	0	0
$\boldsymbol{n_1=7}$	$\boldsymbol{SS_w1=10.00}$		$\boldsymbol{n_2=7}$	$\boldsymbol{SS_w2=10.00}$		$\boldsymbol{n_3=7}$	$\boldsymbol{SS_w3=10.00}$	
$\bar{X}_1=6.00$			$\bar{X}_2=2.00$			$\bar{X}_3=1.00$		

e. Add the SS_w (Sum of Squares Within) for all groups together to get the total SS_w.

$$SS_w = SS_{w1} + SS_{w2} + SS_{w3}$$
$$SS_w = 10.00 + 10.00 + 6.00$$
$$SS_w = 26.00$$

2. Find the df_w by subtracting the number of groups (k) from the total of sample sizes (N).

$$df_w = N - k = 21 - 3 = 18$$
$$df_w = 18$$

3. Find the total SS_b across all groups. The formula for SS_b for each group is as follows:

$$SS_{\text{b\_group}} = n_{\text{group}} \left[\left(\bar{x}_{\text{group}} - \bar{x}_{\text{grand}} \right)^2 \right]$$

We must use this formula to find the SS_b for each group. We will fill in the formula then break down the three steps to using it. If the information is filled in for Group 1, we get the following:

$$SS_{b1} = 7\left[(6.00 - 3.00)^2\right]$$

Now we can use order of operations to find the Sum of Squares Between for Group 1 as follows:

a. Subtract the grand mean ($\bar{x}_{\text{grand}}$) from the mean for Group 1 ($\bar{x}$) to get the group deviation.

$$\bar{x}_1 - \bar{x}_{\text{grand}}$$
$$6.00 - 3.00$$
$$3.00$$

b. Square the group deviation from step 3a.

$$3.00^2 = 9.00$$

c. Multiply the squared deviation for the group (which is the result of step 3b) by the sample size for Group 1 (n_1) to find the Sum of Squares Between for Group 1 (SS_b).

$$SS_{b1} = 7(9.00)$$
$$SS_{b1} = 63.00$$

d. Repeat steps 1a through 1c for each group until the SS_b for each group is known.

Group 2 Calculations	Group 2 Calculations
$SS_{b2} = n_2\left[(barx_2 - \bar{x}_{\text{grand}})^2\right]$	$SS_{b3} = n_3\left[(barx_3 - \bar{x}_{\text{grand}})^2\right]$
$SS_{b2} = 7\left[(2.00 - 3.00)^2\right]$	$SS_{b3} = 7\left[(1.00 - 3.00)^2\right]$
$SS_{b2} = 7\left[(-1.00)^2\right]$	$SS_{b3} = 7\left[(-2.00)^2\right]$
$SS_{b2} = 7(1.00)$	$SS_{b3} = 7(4.00)$
$SS_{b2} = 7.00$	$SS_{b3} = 28.00$

e. Add the SS_b (sum of squares between) for all groups together to get the total SS_b.

$$SS_b = SS_{b1} + SS_{b2} + SS_{b3}$$
$$SS_b = 63.00 + 7.00 + 28.00$$
$$SS_b = 98.00$$

4. Find the df_b by subtracting 1 from the number of groups (k).

$$df_b = k - 1 = 3 - 1 = 2$$
$$df_b = 2$$

5. Write the ANOVA formula with the four values found in the prior steps (i. e. SS_w, df_w, SS_b, and df_b) plugged into their respective locations.

$$F = \frac{SS_b \div df_b}{SS_w \div df_w}$$
$$F = \frac{98.00 \div 2}{26.00 \div 18}$$

6. Solve for MSS_b by dividing SS_b by df_b. This gives you the numerator for the F formula.

$$F = \frac{49.00}{26.00 \div 18}$$

7. Solve for MSS_w by dividing SS_w by df_w. This gives you the denominator for the F formula.

$$F = \frac{49.00}{\mathbf{1.4444}\ldots}$$

8. Finally, divide MSS_b (which is the result of step 6) by MSS_w (which is the result of step 7) to get the obtained F-value.

$$F = \mathbf{33.9230}\ldots$$

The obtained value for this test is 33.92 when rounded to the hundredths place.

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value for F exceeds the critical value as follows: The critical value is 3.555.

The obtained F-value is 33.92.

The obtained F-value exceeds (i.e. is greater than) the critical value. Therefore, the criteria has been met to declare the result significant. This result is significant and supports the hypothesis.

Note

If the hypothesis had a direction, the directions of each group by group comparison would need to be checked in the post-hoc analyses before concluding that the hypothesis had been supported. However, the current hypothesis was simply that groups would differ; thus, we are able to conclude that the hypothesis was supported before checking the post-hoc results. In this scenario, we will still use the post-hoc analyses to determine which group mean(s) were different from which other group mean(s) so that we can provide more detailed information in our APA-formatted summary of the results.

6. **Calculate the effect size and/or other relevant secondary analyses.**

When it is determined that the result is significant, effect sizes should be computed. Because the result was determined to be significant in step 5, the effect size is needed before proceeding to step 7.

The effect size that is appropriate for a one-way ANOVA is a calculation of the percent of variance observed that was systematic (and, thus, was uniquely between groups) which is reported in decimal form. The symbol for this effect size is η^2 (which is named "eta squared"). The formula is as follows:

$$\eta^2 = \frac{SS_b}{SS_T}$$

To calculate this, two things need to be known:

1. the Sum of Squares Between groups (SS_b) and
2. the Sum of Squares Total (SS_T).

We have already found SS_b in earlier steps but still need to find SS_T. SS_T refers to the total sum of squared deviations observed overall for the ANOVA; this is found by summing the SS_b and the SS_w. We already know SS_w from earlier steps. We can summarize the two things we do know and use them to find the one we need as follows:

$$SS_b = 98.00$$
$$SS_w = 26.00$$
$$SS_T = SS_b + SS_w = 98.00 + 26.00 = 124.00$$

Now that we have the pieces needed, we can use this formula to find the proportion of variance that was uniquely accounted for between groups. The calculations for the current data would be as follows:

$$\eta^2 = \frac{98.00}{124.00}$$
$$\eta^2 = 0.7903\ldots$$

Effect sizes, like most values, are rounded and reported to the hundredths place. Thus, this effect size is reported as $\eta^2 = 0.79$.

This translates to 79% of the variance being accounted for by differences between groups rather than within groups. The largest the effect size can be is 1.00 because that would mean 100% of the variance was accounted for by between group differences. In keeping, the lowest an effect size can be is 0.00 because that would mean that 0% of the variance was accounted for by between group differences. Thus, the closer the effect size is to 1.00, the larger it is, and the closer the effect size is to 0.00, the smaller the effect size is. The current effect size of 0.79 would be considered a large effect.

Post-Hoc Analyses for One-Way ANOVA

Comparing the means of three or more independent groups requires two processes:

1. An omnibus test to know whether all means are approximately equal vs. whether at least one mean is significantly different that at least one other mean and,
2. A post-hoc analysis which can specify which means are significantly different than which other means. Post-hoc analyses are warranted when an omnibus result is significant. The omnibus test was significant for Data Set 10.1 and, thus, hos-hoc testing is necessary.

Post-hoc tests for ANOVA are pairwise tests. This means that groups will be compared in pairs (meaning two at a time). The commonly used version of this for data that meet the assumptions for a one-way ANOVA is known as *Tukey's Honestly Significant Difference* (HSD) *post-hoc test*. To use this by hand, an HSD value is calculated and represents the minimum difference that must be observed between any two groups in order to declare the means of the pair significantly different from one another. The HSD formula is as follows:

$$HSD = q\sqrt{\frac{MS_w}{n}}$$

Three things are needed to find an $HSD: MS_w$, n, and q. We already solved for two of those during omnibus testing but still must find q. In this formula, q is known as the studentized range statistic. We can locate q using the table in Appendix F. To find q, we need to know k and df_w. Using these, we can find \(q = 3.609\) in Appendix F.

$$HSD = 3.609\sqrt{\frac{1.4444\ldots}{7}} = 1.6394\ldots \approx 1.64$$

Then, the pairwise differences must be calculated to compare them to the HSD threshold. The pairwise differences in means for Data Set 10.1 are as follows:

Descriptive Statistics by Group		
Group 1: M = 6.00, SD = 1.29	Group 2: M = 2.00, SD = 1.29	Group 3: M = 1.00, SD = 1.00
Pairwise Comparisons		
Group 1 vs. Group 2	Group 1 vs. Group 3	Group 2 vs. Group 3
6.00 – 2.00 = 4.00	6.00 – 1.00 = 5.00	2.00 – 1.00 = 1.00
4.00 > 1.64, p < .05	5.00 > 1.64, p < .05	1.00 < 1.64, p > .05

The pairwise comparisons indicate that the means of Group 1 and Group 2 are significantly different, the means of Group 1 and Group 3 are significantly different, but that the means of Group 2 and Group 3 are not significantly different from one another.

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

APA format requires a specific organization be used for reporting the results of a test. This includes a specific format for reporting relevant symbols and details for the formula and data used. Following this, the results for our hypothesis with Data Set 10.1 can be written as shown in the summary example below.

APA Formatted Summary Example

A one-way ANOVA was used to test the hypothesis that the mean acts of aggression would be different among children in three different conditions. Consistent with the hypothesis, the mean acts of aggression were different among the conditions, $F(2, 18) = 33.92$, p < .05. The effect size of 0.79 was large. Tukey's HSD post-hoc results indicate that the mean acts of aggression was significantly higher in group exposed to aggressive acts (M = 6.00; SD = 1.29) than the group exposed to neutral acts (M = 2.00; SD = 1.29) and the group which was not exposed to either (M = 1.00; SD = 1.00), p < .05. However, the means were not significantly different between Group 2 and Group 3, p > .05.

As always, the APA-formatted summary provides a lot of detail in a particular order. For a brief review of the structure for the APA-formatted summary of the omnibus test results, see the summary below.

Summary of APA-Formatted Results for the One-Way ANOVA

In your APA write up for a one-way ANOVA you should state:

1. Which test was used and the hypothesis which warranted its use.
2. Whether the aforementioned hypothesis was supported or not. To do so properly, three components must be reported:
   a. The mean and standard deviation for each group
   b. The test results in an evidence string as follows: $F\left(df_b, df_w\right)$ = obtained value
   c. The significance portion of the evidence string for the omnibus test as p < .05 if significant or p > .05, ns if not significant
   d. The effect size (Note: this part is only required if the omnibus result was significant).
   e. The significance for the post-hoc pairwise comparisons as p < .05 if significant or p > .05, ns if not significant (Note: this part is only required if the omnibus result was significant).

Anatomy of the Evidence String

The following breaks down what each part represents in the evidence string for the omnibus results in the APA-formatted paragraph above:

Symbol for the test	Degrees of Freedom	Obtained Value	p-Value
F	(2, 18)	= 33.92,	p < .05.

Reading Review 10.3

1. How is SS_b calculated?
2. What two things are reported inside the parentheses of the evidence string?
3. What information is needed to find the critical value for a one-way ANOVA?
4. Which two things only need to be included in the results summary paragraph when the omnibus results are significant?
5. How can an effects size be calculated for a one-way ANOVA?

10.5: Addressing Violations to Assumptions

As noted earlier in this chapter, there are several assumptions to the one-way ANOVA. When they are met, we can proceed using the standard procedures described throughout this chapter. However, when assumptions of tests are not met it means either the test cannot be used or that a modified or alternative formula must be used. Most of the assumptions have to do with the types of variables and measures used and those cannot be violated. You must always have two or more independent groups measured on the same, quantitative variable. However, it is possible to proceed with the one-way ANOVA if the other assumptions are not all met. Those assumptions include approximate normality for the dependent variable, no problematic outliers, homogeneity of sample sizes, and homogeneity of variances.

Thankfully, the one-way ANOVA is fairly robust to violations, especially when the sample sizes are large and even. Nevertheless, it is good practice to check the homogeneity of variances. As was true of the independent samples *t*-test, homogeneity of variances can be checked for a one-way ANOVA using the Levene's test. This is often done with the aid of SPSS software. When variances are not homogeneous, a Welch's ANOVA with a Games-Howell post-hoc test should be considered instead of a standard one-way ANOVA with a Tukey's post-hoc test.

When conducting a one-way ANOVA in SPSS, users can choose to include a Levene's test. With this in mind, let's turn to how to conduct a one-way ANOVA using SPSS, paying attention to when and how to check the assumption of homogeneity of variances.

10.6: Using SPSS

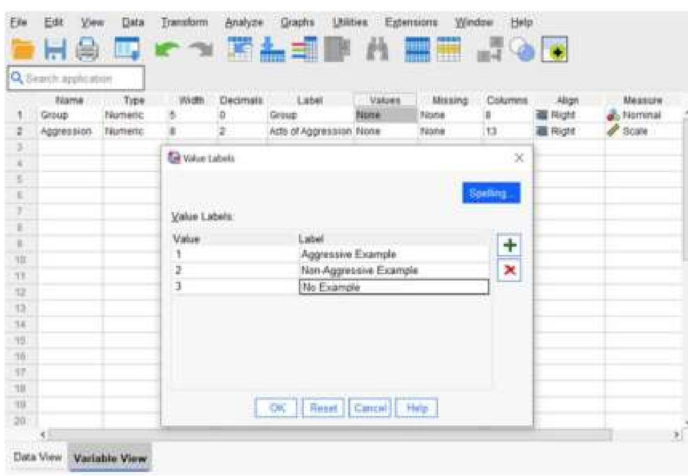
As reviewed in Chapter 2, software such as SPSS can be used to expedite analyses once data have been properly entered into the program. This section focuses on how to enter and analyze data for a one-way ANOVA using SPSS. SPSS version 29 was used for this book; if you are using a different version, you may see some variation from what is shown here.

Entering Data

The one-way ANOVA is bivariate. One variable is used to organize data into comparison groups. The other variable is being compared between those groups. The variable being compared must be quantitative and should have been measured using numbers on an interval or ratio scale. If these things are all true of your data, you are ready to open SPSS and begin entering your data.

Open the SPSS software, click “New Dataset,” then click “Open” (or “OK” depending on which is shown in the version of the software you are using). This will create a new blank spreadsheet into which you can enter data. Click on the Variable View tab on the bottom of the spreadsheet. This tab of the spreadsheet has several columns to organize information about the variables. The first column is titled “Name.” Start here and follow these steps:

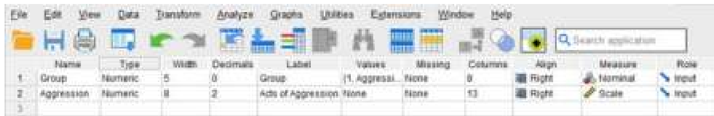
1. Click the first cell of that column and enter the name of your grouping variable using no spaces, special characters, or symbols. You can name this variable “Group” for simplicity. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
2. Click the first cell of the column titled “Type” and then click the three dots that appear in the right side of the cell. Specify that the data for that variable appear as numbers by selecting “Numeric.” For numeric data SPSS will automatically allow you to enter values that are up to 8 digits in length with decimals shown to the hundredths place as noted in the “Width” and “Decimal” column headers, respectively. You can edit these as needed to fit your data, though these settings will be appropriate for most variables in the behavioral sciences.
3. Click the first cell of the column titled “Label.” This is where you can specify what you want the variable to be called in output, including in tables and graphs. You can use spaces or phrases here, as desired.
4. Click on the three dots in the first cell of the column titled “Values.” This is where you can add details about each group. Click the plus sign and specify that the value 1 (for Group 1) refers to the Aggressive Example subsample. Then click the plus sign again and specify that value 2 (for Group 2) refers to the Non-Aggressive Example subsample as shown below. Then click the plus sign again and specify that value 3 (for Group 3) refers to the No Example subsample as shown below. Then click “OK.”



5. Click on the first cell of the column titled “Measure.” A pulldown menu with three options will allow you to specify the scale of measurement for the variable. Select the “Nominal.” option because grouping variables are nominal. Now SPSS is set-up for data for the grouping variable.
6. Next we need to set up space for the quantitative variable. In the second cell (row) of the “Name” column, enter the name of your quantitative variable using no spaces, special characters, or symbols. You can name this variable “Aggression” for simplicity. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
7. Click the cell in row 2 of the column titled “Label.” Here we can clarify that Aggression refers to “Acts of Aggression” by stating as such in the label column for this variable.

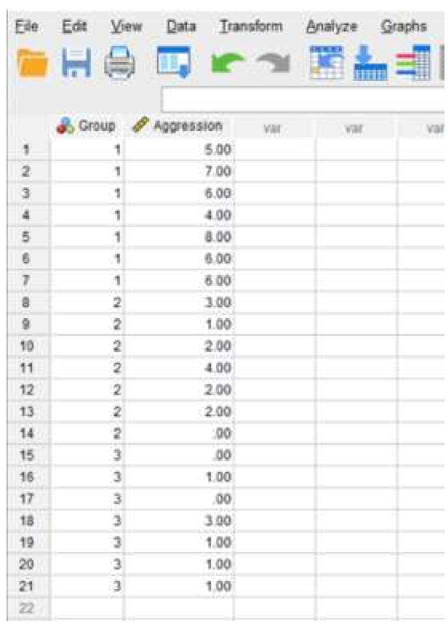
8. Click the cell in row 2 of the column titled "Type" and then click the three dots that appear in the right side of the cell. Specify that the data for that variable appear as numbers by selecting "Numeric." Again, you can edit the width and decimals as needed to fit your data.
9. Click on the cell in the second row of the column titled "Measure." A pulldown menu with three options will allow you to specify the scale of measurement for the variable. SPSS does not differentiate between interval and ratio and, instead, refers to both of these as "Scale." Select the "Scale" option because if you are using a one-way ANOVA your data for this variable should have been measured on the interval or ratio scale.

Here is what the Variable View tab would look like when created for Data Set 10.1:



Now you are ready to enter your data. Click on the Data View tab toward the bottom of the spreadsheet. This tab of the spreadsheet has several columns into which you can enter the data for each variable. Each column will show the names given to the variables that were entered previously using the Variable View tab. Click the first cell corresponding to the first row of the first column. Start here and follow these steps:

1. Enter the data for the grouping variable moving down the rows under the first column. Put a 1 in this column for everyone who is a member of Group 1, put a 2 for this column for everyone who is a member of Group 2, and put a 3 for this column for everyone who is a member of Group 3. Continue in this fashion if you have more than three groups until all data are entered for the grouping variable.
2. Enter the data for the quantitative variable moving down the rows under the second column. If your data are already on your computer in a spreadsheet format such as excel, you can copy-paste the data in for the variable. Take special care to ensure the Aggression data for Group 1 appear in the rows corresponding to Group 1, that the Aggression data for Group 2 appear in the rows corresponding to Group 2, and that the Aggression data for Group 3 appear in the rows corresponding to Group 3.
3. Then hit save to ensure your data set will be available for you in the future.

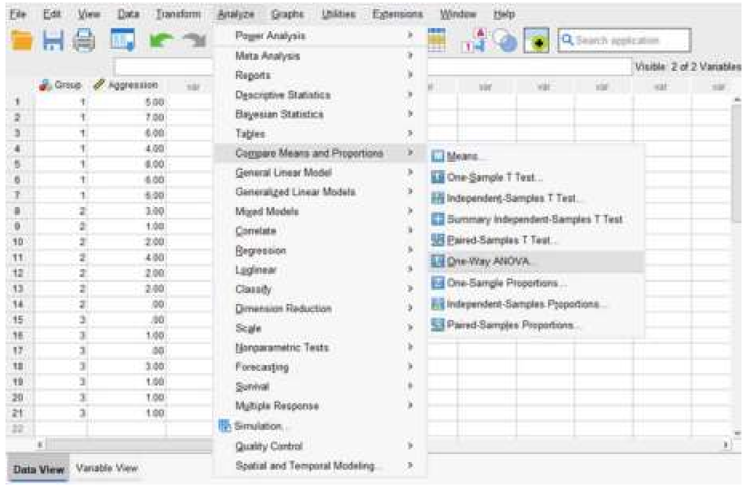


Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

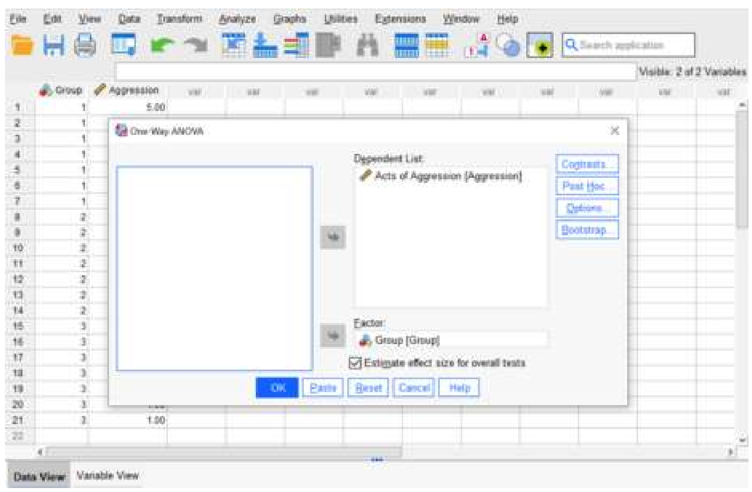
Conducting a one-way ANOVA in SPSS

The steps to running a one-way ANOVA in SPSS are:

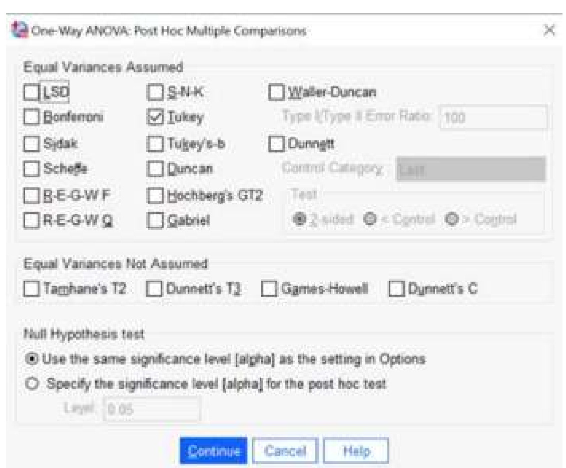
1. Click Analyze -> Compare Means and Proportions -> One-Way ANOVA from the pull down menus as shown below.



2. Drag the name of the quantitative variable from the list on the left into the Dependent list box on the right of the command window. You can also do this by clicking on the variable name to highlight it and the clicking the arrow to move to the desired location. Next, put the grouping variable into the Factor box on the right side of the command window. If the version of SPSS you are using has a check box to estimate effect sizes, click that as well.



3. Click the Options tab. Select both “Descriptive” and “Homogeneity of variance test.” Then click “Continue.”
4. Click the Post Hoc tab. Once in that section, select Tukey Then click “Continue.”



5. Click “OK” to run the analyses.

6. The output (which means the page of calculated results) will appear in a new window of SPSS known as an output viewer. The results will appear in six tables as shown below.

Descriptives

Acts of Aggression					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Aggressive Example	7	6.0000	1.29099	.48795	4.8060	7.1940	4.00	8.00
Non-Aggressive Example	7	2.0000	1.29099	.48795	.8060	3.1940	.00	4.00
No Example	7	1.0000	1.00000	.37796	.0752	1.9248	.00	3.00
Total	21	3.0000	2.48998	.54336	1.8666	4.1334	.00	8.00

Tests of Homogeneity of Variances

Acts of Aggression	Levene Statistic	df1	df2	Sig.
Based on Mean	.255	2	18	.777
Based on Median	.255	2	18	.777
Based on Median and with adjusted df	.255	2	17.743	.777
Based on trimmed mean	.217	2	18	.807

ANOVA

Acts of Aggression	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	98.000	2	49.000	33.923	<.001
Within Groups	26.000	18	1.444		
Total	124.000	20			

ANOVA Effect Sizes[a]

Acts of Aggression		95% Confidence Interval	
	Point Estimate	**Lower**	**Upper**
Eta-squared	.790	.522	.862
Epsilon-squared	.767	.469	.847
Omega-squared Fixed-effect	.758	.457	.840
Omega-squared Random-effect	.611	.296	.725
a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.			

Multiple Comparisons

Dependent Variable: Tukey HSD	Acts of Aggression				95% Confidence Interval for Mean	
(I) Group	**(J) Group**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**Lower Bound**	**Upper Bound**
Aggressive Example	Non-Aggressive Example	4.00000*	.64242	<.001	2.3604	5.6396
	No Example	5.00000*	.64242	<.001	3.3604	6.6396
Non-Aggressive Example	Aggressive Example	-4.00000*	.64242	<.001	-5.6396	-2.3604
	No Example	1.00000*	.64242	.289	-.6396	2.6396
No Example	Aggressive Example	-5.00000*	.64242	<.001	-6.6396	-3.3604
	Non-Aggressive Example	1.00000*	.64242	.289	-2.6396	.6396
*. The mean difference is significant at the 0.05 level						

Acts of Aggression

Tukey HSD[a]		Subset for alpha = 0.05	
	N	**1**	**2**
No Example	7	1.0000	
Non-Aggressive Example	7	2.0000	
Aggressive Example	7		6.0000
Sig.		.289	1.000
Means for groups in homogeneous subsets are displayed. a. Uses Harmonic Mean Sample Size = 7.000.			

Reading SPSS Output for One-Way ANOVA

The first table shows the descriptive statistics for the test. These include several statistics such as the sample sizes, means, standard deviations, and the standard errors. These match the results from the hand-calculations performed earlier in this chapter for Data Set 10.1. The means and standard deviations are needed for summarizing the results for each group in an APA-formatted summary paragraph.

The second table shows one of the assumptions checks: homogeneity of variances. The Levene's test for homogeneity based on means should be checked in this table. When variances are homogeneous enough to meet this assumption, the Levene's test will have a non-significant *p*-value (meaning that the group variances are not significantly uneven). In the output we see the Levene's test has a "Sig." value (which is what SPSS calls the *p*-value) of .777. This *p*-value is greater than .05 indicating that the group variances are *not* significantly uneven. This is desirable and means we have met the assumption of homogeneity of variances and can proceed to reading our results for the ANOVA.

The third table shows the main test results which are needed for the evidence string, including the F-value, the degrees of freedom between (df_b), the degrees of freedom within (df_b), and the *p*-value for the omnibus test. This table is sometimes referred to as a "Source Table" because it summarizes all the main parts of the ANOVA formula, the result of the formula, and the significance. We can see that the sums of squares, degrees of freedom, mean sums of squares, and the F-value all match the hand-calculations we performed earlier in this chapter using Data Set 10.1. This also includes the information needed to create the second sentence of the APA-formatted results, including the values needed for its evidence string.

The fourth table provides the effect sizes. The one of focus for this chapter and Data Set 10.1 appears in the first row of the table labelled as "Eta-squared." The value of the effect size appears in decimal form under the column titled "Point Estimate." Here we see the effect size is .79 when rounded to the hundredths place; this matches what was found using hand calculations earlier in this chapter.

The fifth and sixth tables provide the results for the post-hoc tests in two formats. The fifth table provides more details and the sixth reiterates some of those details in a different format to accentuate group similarity and dissimilarity. Let's first look at the fifth table which shows the multiple comparisons. Each pairwise comparison is shown in the fifth table twice in the following order:

Group 1 vs. Group 2

Group 1 vs. Group 3

Group 2 vs. Group 1

Group 2 vs. Group 3

Group 3 vs. Group 1

Group 3 vs. Group 2

You only need to review three of these because the other three are redundant to them (e.g. Comparing Group 1 to Group 2 is the same as comparing Group 2 to Group 1). SPSS will call the first group in each of these pairs "I" and the second in each pair "J." The two most important columns to review are the "Mean Difference" column and the "Sig." column. The mean difference column subtracts the mean of the second group named in the pair (called the "J Group") from the mean of the first group named in the pair (called the "I Group."). The larger the value is in the mean difference column, the greater the difference is in the means before accounting for error. The next column to check, and the most useful, is the "Sig." column. This is where we will see the p-value for each pairwise comparison. Keep in mind that p refers to the probability of a Type I Error and that a result is significant when this value is less than .05. Thus, for each pair, if the p-value shown is less than .05, it indicates the means of the two groups are significantly different. With this in mind, we can interpret the results in the table.

Let's look at the comparison of Group 1 (Aggressive Example Group) to Group 2 (Non Aggressive Example Group) as an example. The difference in the means of those two groups was 4.00 (when shown to the hundredths place). This indicates that the mean of Group 1 was 4.00 units higher than the mean of Group 2. Now we can look at the "Sig." column to find the p value. It shows that the p-value is so small that it is less than .001. This indicates that the chance of a Type I Error is very small. Thus, after accounting for error, the difference in the means of these two groups is found to be statistically significant because the p-value is less than .05.

Multiple Comparisons

Dependent Variable: Tukey HSD	Acts of Aggression				95% Confidence Interval for Mean	
(I) Group	**(J) Group**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**Lower Bound**	**Upper Bound**
Aggressive Example	Non-Aggressive Example	4.00000*	.64242	<.001	2.3604	5.6396
	No Example	5.00000*	.64242	<.001	3.3604	6.6396
Non-Aggressive Example	Aggressive Example	-4.00000*	.64242	<.001	-5.6396	-2.3604
	No Example	1.00000*	.64242	.289	-.6396	2.6396
No Example	Aggressive Example	-5.00000*	.64242	<.001	-6.6396	-3.3604
	Non-Aggressive Example	1.00000*	.64242	.289	-2.6396	.6396

Dependent Variable: Tukey HSD	Acts of Aggression		95% Confidence Interval for Mean
*. The mean difference is significant at the 0.05 level			

Following the same logic, we can check the other two pairwise comparisons. We can see that the difference in the means for Group 1 (Aggressive Example Group) and Group 3 (No Example Group) is 5.00 (when shown to the hundredths place). This means the mean of Group 1 was 5.00 units higher than the mean of Group 3. After accounting for error, this is found to be a statistically significant difference because the p-value is less than .05. Finally, let's compare Group 2 (Non-Aggressive Example Group) to Group 3 (No Example Group). The difference in the means of these two groups is 1.00 (when shown to the hundredths place) which is noticeably smaller than in the other two pairwise comparisons. The p-value for this pair is .289, which is higher than .05. Taken together, the difference in the means is small and the chance of a Type I Error if we conclude these two groups are different is unacceptably high. Thus, the difference between Group 2 and Group 3 is *not* statistically significant.

The sixth table displays the main results from the multiple comparison table in a new format. The sixth table shows the name of the quantitative variable being compared between groups as its title. For Data Set 10.1 this is "Acts of Aggression." It then provides summaries of each group in their own row. The summaries include just the sample size and mean for each group. However, it will put the means into columns that distinguish which means are not significantly different (by putting them in the same column) and which are different (by putting them in separate columns). Here we see that the means for the No Example Group (Group 3) and the Non-Aggressive Example Group (Group 2) are in the same column. This indicates that they are not significantly different from one another. However, the mean for the Aggressive Example Group (Group 1) appears in a column separate from the other group means. This indicates that the mean for this group is significantly different than the means for the other two groups. Some people may find this version easier to understand but you can use either the fifth or sixth table to check which groups are significantly different in the post-hoc analyses. Because these two tables show the same key results in two different ways, you only need to use the one which you prefer when checking and reporting your post-hoc results.

Acts of Aggression

Tukey HSD[a]		Subset for alpha = 0.05	
	N	**1**	**2**
No Example	7	1.0000	
Non-Aggressive Example	7	2.0000	
Aggressive Example	7		6.0000
Sig.		.289	1.000
Means for groups in homogeneous subsets are displayed. a. Uses Harmonic Mean Sample Size = 7.000.			

These results are all consistent with what we found following the hand-calculations earlier in this chapter, which is as expected. The benefit to doing the calculations by hand is that we get clarity on what goes into each formula, why, and how it connects to the result. The benefits to using SPSS, of course, are that it is fast and easy to use. However, we must always keep in mind that SPSS cannot think for us. Instead, it just computes what we tell it to. It is up to us to know when to use the formula, how to check that the assumptions are met, to ensure the data are entered properly, and finally, to interpret the results appropriately. With this in mind, let's consider these results as we would if they were part of a real study of aggression.

Real-World Interpretations of One-Way ANOVA

Through this book and our classes, we are learning how to use statistics as a tool to measure, test, and ultimately better understand truths about the world. In some behavioral and social sciences, like psychology, statistics is used to test hypotheses about the experiences and behaviors of humans. In keeping, this chapter focused on an example about whether what children were exposed to might impact their aggressive behaviors. If this was done as a true experiment where children were randomly assigned to one of three different conditions (i.e. exposure groups) and their behavior was measured and compared across those conditions, a one-way ANOVA could be used as a tool to test whether exposure impacted behavior.

In the example in this chapter (which used fake data made up for demonstration and practice purposes), it was hypothesized that children who were shown an adult acting aggressively toward a toy (Group 1), children who were shown an adult playing non-aggressively with the toy (Group 2), and children who were not shown any interactions with it (Group 3) would engage in different mean acts of aggression toward the toy themselves. The results can be interpreted to say that the children who were shown the aggressive example engaged in more aggressive behaviors, on average, than those who were shown a non-aggressive example and those who were shown no example. However, the amounts of aggressive behaviors engaged in by children shown the non-aggressive example were no different, on average, than those shown no example. Were these results obtained from a real study, it could be compelling evidence for a theory stating that people learn from, and replicate, behaviors that they observe in others.

Though our example is fake, it is based on real research. Such a theory does exist and is known as the theory of Observational Learning. This was famously tested by Albert Bandura, Dorothea Ross, and Sheila A. Ross in 1961. Their original study (known by many as the Bobo Doll Experiment) is often presented as having only three groups, much like the example in this chapter. However, the original study also grouped based on the gender of the child and the gender of the adult observed to assess any gender-related differences. In addition, the data in the original study did not have homogenous variances. For these reasons, the researchers used a Friedman two-way analysis of variance rather than the one-way ANOVA reviewed in this chapter.

Some of the things we should learn from this chapter and this example are:

1. Inferential tests such as ANOVA are tools that, when understood and used appropriately, can help us learn new things about our world and
2. Real research is often more complicated and messy (i.e. assumptions may be violated necessitating adjustments to the analyses or formulas used) than we would like it to be.

Thus, each new statistical skill we acquire opens up another way of being able to test and understand our world. The examples we learn in class give us a solid foundation onto which we can continue to acquire more tools to deal with the various and complicated aspects of our world.

Reading Review 10.4

1. What scale of measurement should be indicated in SPSS for the grouping (factor) variable?
2. What information is used in the output to check that the assumption of homogeneity of variances was met?
3. Under which table and column of the SPSS output can the F-value be found?
4. Under which table and column of the SPSS output can the omnibus p-value be found?
5. Under which table and column of the SPSS output can the η^2-value be found?
6. Under which table and column of the SPSS output can the post-hoc p-values be found?

10.7: Structured Summary for the One-Way ANOVA

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes.

Summarize what each symbol stands for, assuming three groups are being compared.

n_1 =

n_2 =

n_3 =

N =

$\bar{X}_1$ =

$\bar{X}_2$ =

$\bar{X}_3$ =

$\bar{X}_{\text{grand}}$ =

s_1 =

s_2 =

s_3 =

SS_{w1} =

SS_{w2} =

SS_{w3} =

SS_w =

SS_{b1} =

SS_{b2} =

SS_{b3} =

SS_b =

df_w =

df_b =

MSS_w =

MSS_b =

Fill-in the appropriate information for each section below:

1. One-way ANOVA Basics
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What assumptions must the data meet to use this test?
2. One-way ANOVA Formula
   a. What is the formula for a one-way ANOVA?
   b. What things should be computed in the preparatory steps for using this formula?
   c. What are the steps for solving using this formula?
3. Reporting Results from a one-way ANOVA
   a. How is this statistic reported when using APA format?
      i. What things must be reported in the APA summary sentence for the omnibus test?

ii. What two specific statistics must be reported for each of the independent groups (which may appear as part of the post-hoc results sentences)?
iii. How is effect size computed and reported for a one-way ANOVA?
iv. How are post-hoc comparisons computed and reported for a one-way ANOVA?

10.8: One-Way ANOVA Calculations Chart

Phase 1: Calculations For Each Group						
Find the five statistics that focus on calculations for each group, one by one.*		**Sample Sizes**	**Means**	**Sum of Squares Within**	**Standard Deviations**	**Sum of Squares Between**
	Group 1	n_1 =	$\bar{X}_1$=	SS_{w1}=	s_1 =	SS_{b1} =
	Group 2	n_2 =	$\bar{X}_2$=	SS_{w2}=	s_2 =	SS_{b2} =
	Group 3	n_3 =	$\bar{X}_3$=	SS_{w3}=	s_3 =	SS_{b3} =
Phase 2: Summary Calculations						
Find the five statistics that use data across groups.	**Summary Components**	N =	$\bar{X}_{\text{grand}}$ =	SS_w =	k =	SS_b =
Find both forms of degrees of freedom.	**Degrees of Freedom**			df_w =		df_b =
Phase 3: ANOVA (omnibus) Calculations						
Plug the parts into the ANOVA formula and find F.	**Between**	SS_b =	df_b =	MSS_b =	F =	Critical Value =
	Within	SS_w =	df_w =	MSS_w =	Is the result significant?	
Phase 4: Post-ANOVA Calculations						
If ANOVA is significant, complete follow-up calculations.	**Follow-Up**	**Formula**	**How to use and interpret it:**			
	Effect Size	$\eta^2=\frac{SS_b}{SS_T}$				
	Post-Hoc (Tukey)	$HSD=q\sqrt{\frac{MS_w}{n}}$				

**Note*: The first section shows room for three groups, however, rows can be added to accommodate the data when using ANOVA to compare more than three groups.

CHAPTER OVERVIEW

11: Repeated-Measures ANOVA

11.1: Introduction to Repeated-Measures ANOVA

The **repeated-measures ANOVA** is used when you want to test whether the means of a single group are significantly different when measured at different exposures to the independent variable or waves of testing. A wave refers to a time point, usually connected to some kind of intervention or other change in an independent variable. Waves may also be called times, conditions, or treatments. Recall that ANOVA is short for analysis of variance because the formula examines and compares the proportions of variability attributed to different sources. If the means are distinct enough at the different waves of measurement, even after accounting for the fact that there is some variability around the mean for each wave of testing, the result will be significant.

Repeated-measures ANOVA is a bivariate technique. One variable is a qualitative grouping variable where data from a single sample are grouped based on different times, conditions, and/or exposures to the independent variable. The other is a comparison or focal quantitative variable. This quantitative variable is measured at each wave or condition using the same group of participants. The qualitative grouping variable is often referred to as a **factor.** Thus, factors refer to how the independent variables or conditions are grouped. Data from the sample in each of those conditions are compared.

Comparing ANOVAs

In repeated-measures ANOVA, the quantitative variable is measured in the same group of participants under different conditions of the qualitative variable. This is different from independent-groups ANOVA where data from different groups of participants are compared for each condition (see Chapter 10 to review independent-groups ANOVA).

There are a few reasons a repeated-measures ANOVA, and a corresponding research design, can be useful. First, repeated-measures ANOVA is similar to a dependent samples *t*-test but without the limitation of only being able to compare two groups (waves) of data (see Chapter 8 to review the dependent samples *t*-test). Second, repeating measures within a single group can reduce or eliminate cohort or pre-existing grouping effects that may occur when comparing several independent groups. Cohort and pre-existing grouping effects refer to when there are differences in the groups that may cause them to have different mean scores of the quantitative outcome variable that are not due to differences in the qualitative grouping (factor) variable. Another way of saying this is that the main issue that can arise when you put multiple, independent samples into different experimental conditions, is that you cannot always be sure whether differences observed in those groups are due to pre-existing differences in those groups or due to the different conditions they experienced as part of the study. Using the same group across conditions can be used to address this issue.

A **confound** or **confounding variable** refers to something that erroneously causes it to appear that the groups differed because of the study conditions when, in fact, the differences were not due to those conditions**.** When different participants are compared for each condition, pre existing differences among members of the different groups are possible confounds. To avoid this, therefore, some studies are designed to measure the same group multiple times under different conditions. These are known as a repeated-measures design studies. When the same group is being compared on the same quantitative outcome variable across three or more conditions, a repeated-measures ANOVA is the appropriate analysis. However, it should be noted that there are other possible issues with repeated-measures designs such as learning effects. The details about the benefits and limitations of research designs, however, is beyond the scope of this book so we will briefly differentiating between two aspects of research methods: 1. independent-groups designs vs. repeated-measures designs and 2. experimental vs. non-experimental designs because these have implications for how we analyze the data and whether we can make statements of cause-effect from the results of those analyses. However, further aspects of research methods are covered in detail in courses on Research Methods which are common in social and behavioral science programs.

11.2: Variables, Data, and Hypotheses that Fit a Repeated-Measures ANOVA

Variables

The repeated-measures ANOVA requires the use of one qualitative factor or treatment variable. Each of these terms simply refers to the fact that different conditions of one variable are being experienced by the sample. This type of factor variable is the independent variable (IV) when a true experiment is being performed. However, repeated-measures ANOVA can also be used with non-experimental designs in which case this variable is more appropriately referred to as a factor or grouping variable rather than a true IV.

The other variable is quantitative and is measured the same way under each of the conditions; this variable is the dependent variable (DV) when a true experiment is performed. This quantitative variable is the thing being compared under each of the conditions. For example, a group of individuals could be given four meals, each prepared using different versions of a recipe, and asked to rate how much they enjoyed each meal on a scale of 0 to 10. In this example, the individuals make up the sample and would report their enjoyment four times each, once for each of the meals. The different versions of the meal make up the different conditions or factors. The ratings of enjoyment make up the focal outcome variable (DV). When ANOVA is used with a non experimental design, the DV is sometimes referred to as a “test variable” to clarify that it is not a true DV.

Data

Each statistical test has some assumptions which must be met in order for the formula to function properly. In keeping, there are a few assumptions about the data which must be met before a repeated-measures ANOVA is used. Most of these are the same that must be true for dependent samples *t*-tests. First, the data for the same continuous, quantitative variable must have been measured multiple times on the same interval or ratio scale. Second, the scores from all of the waves of measurement must be matched. This means that the same participants have scores for each condition that can be identified together such that the researcher knows which score at each wave belongs to each person. A researcher needs to be able to identify the scores for each person so they can see how much that participant’s score changed from one wave of testing to another. Third, data for the quantitative variable should be fairly normally distributed in each group without notable impact due to outliers (such as problematic skew). Finally, there is homogeneity of variances but extended to comparisons between more than just two waves of data.

Hypotheses

Hypotheses for the repeated-measures ANOVA must include both the qualitative factor variable (which may also be referred to as the treatment, grouping, or independent variable) and a quantitative variable (which may also be referred to as the test, outcome, or dependent variable). ANOVA uses a one-tailed test which is often referred to as a right-tailed test (referring to the right side of a distribution). Therefore, we will review the different constructions of hypotheses which are appropriate for ANOVA but those will not be used to differentiate between using a one-tailed or a two-tailed test.

Just like for a one-way ANOVA, repeated-measures ANOVA is broken down into two tests: an omnibus test and a post-hoc test. The omnibus test is used to test whether the mean measured at any one time is significantly different from the mean measured at another time. However, the omnibus test cannot tell you which means were different from which other means. This is why a post-hoc test is needed when an omnibus test result is significant. The post-hoc test is used to check each group mean against the others to determine which ones were significantly different than each other, two means at a time. Comparisons done two at a time are often referred to as “pairwise comparisons.” Thus, we can think of the omnibus portion of the ANOVA as useful in assessing a non-directional hypothesis by testing whether any means were different and the post-hoc portion as useful in assessing any directional aspects of the hypothesis. Keep these aspects of the repeated-measures ANOVA in mind as we review the hypotheses.

For the repeated-measures ANOVA, the non-directional research hypothesis is that the means are not the same under each condition or wave of testing. This would mean that at least one mean would be different than at least one other mean, breaking a pattern of equality across all means. The corresponding null hypothesis is that the means are the same under each condition or wave of testing. Note that because there can be more than three conditions in an ANOVA, only three will be shown and ellipses will be used to indicate that the hypothesis could be expanded to include more than just those three. The non-directional research and corresponding null hypotheses can be summarized as follows:

Non-Directional Hypothesis for a Repeated-Measures ANOVA

Research hypothesis	The means for each condition or wave of testing are not all equal to each other.	$H_A : \mu_1 \neq \mu_2 \neq \mu_{3...}$

Null hypothesis	The means for each condition or wave of testing are all equal to each other.	$H_0 : \mu_1 = \mu_2 = \mu_3 \ldots$

There are many different directional hypotheses possible for the repeated-measures ANOVA. We will review just one of them as an example presuming there are three groups being compared. One possible directional research hypothesis is that the mean for Condition 1 will be *greater than* the mean for Condition 2 and that the mean for Condition 2 will be *greater than* the mean for Condition 3. The corresponding null hypothesis is that the means for Conditions 1 through 3 will be less than or equal to each other, respectively. This version of the research and corresponding null hypotheses can be summarized as follows:

Example of a Directional Hypothesis for a Repeated-Measures ANOVA

Research hypothesis	The mean of Condition 1 will be greater than the mean of Condition 2 and the mean for Condition 2 will be greater than the mean for Condition 3.	$H_A : \mu_1 \, \mu_2 > \mu_{3\ldots}$
Null hypothesis	The mean of Condition 1 will not be greater than the mean of Condition 2 and/or the mean for Condition 2 will not be greater than the mean for Condition 3.	$H_0 : \mu_1 \leq \mu_2 \leq \mu_3 \ldots$

Note that testing a non-directional hypothesis (which will be the focus of this chapter) only requires the use of the omnibus test but that a directional hypothesis would require the use of both the omnibus test and post-hoc analyses. A non-directional hypothesis is supported when the repeated-measures ANOVA omnibus result is significant. A directional hypothesis is supported when the both the omnibus result is significant and the post-hoc results are in the direction hypothesized and also significant.

Experimental Design and Cause-Effect.

The repeated-measures ANOVA is sometimes used to analyze data from an experiment when multiple conditions are being tested and compared. However, it can also be used to test data collected using non-experimental research methods. When an experimental design is used and other features are present, it may be appropriate to use causal language when interpreting and/or reporting results. For example, the grouping variable can be referred to as a true independent variable (IV) or as an experimental variable when an experiment is performed. It can also be appropriate to report whether different conditions of the IV caused changes in the dependent variable. However, when non-experimental methods are used causal language is not appropriate. In these cases, it is more accurate to use terms like grouping variable, factor, or condition rather than IV or causal variable. It is best to use non-causal language as a default and to only switch to using causal language when it is known that an experimental design was used and that causal language is appropriate (see Chapter 8 for a review of experimental designs and causal language).

What is a Factor?

In a repeated-measures design, the qualitative variable refers to the conditions, factors, treatments, or waves the same participants experienced. This variable is often called a factor variable in non-experimental research and an independent variable in experimental research. The quantitative variable is measured under each condition using the same participants. This is often referred to as the test variable or focal variable in non-experimental research and the dependent variable in experimental research.

Reading Review 11.1

1. What assumptions must be met before using a repeated-measures ANOVA?
2. What is a non-directional hypothesis that could be tested using a repeated-measures ANOVA?
3. Which terms may be used to refer to the qualitative variable when using repeated measures ANOVA?
4. Which terms may be used to refer to the quantitative variable when using repeated measures ANOVA?

11.3: The Foundations of Repeated-Measures ANOVA Formula

The omnibus test is performed using the ANOVA F-formula. When we test a hypothesis using repeated measures ANOVA, it is because we expect that a substantial proportion of variability will be attributed to the conditions relative to how much is attributed to random error. Therefore, the focus of the computations is how much variation is attributed to the conditions relative to how much is attributed to random error. The repeated measures ANOVA formula uses sums of squares to measure variability. Though the focus is on two sources, the variability can come from three important sources: different conditions, differences among participants, and random error. Therefore, the sums of squares must be parsed out and attributed to these three different sources so that the proportions attributed to conditions vs. random error can be compared.

The formula for repeated measures ANOVA can be written as follows:

$$F=\frac{MSS_{\text{between}}}{MSS_{\text{error}}}$$

The formula can be understood as follows:

$$F=\frac{\text{variation between conditions}}{\text{error variance}}$$

This is similar to the independent groups ANOVA (see Chapter 10) but with one very important difference: the denominator for repeated measures ANOVA is a computation of random error instead of the sum of squares within (SS_w). The numerator for repeated-measures ANOVA focuses on differences between conditions and the denominator focuses on random, unsystematic differences observed, also known as the error variance.

The obtained value, F, tells us the ratio of difference between conditions relative to random error. Another way to say this is that it tells us how different the means for conditions are from one another after removing individual variation that is consistent and taking into account random variation that is not attributed to the conditions nor to the individuals. The goal is to see how much of the difference observed is attributed to the conditions relative to how much is attributed to random, unsystematic variation alone. The variation of interest between conditions can be thought of as systematic differences caused by the independent variable (i.e. the conditions) rather than individual differences (i.e. individual variation) or random error. The portion of variation between and attributed to the conditions is known as the **treatment effects.** However, when we observe variation between conditions, it will include not only the treatment effects but also some random, unsystematic variation that can occur known as **random error.** Because both treatment effects and random error exist between groups, random error must be accounted for in the denominator of the ANOVA formula. To summarize, we can understand the formula's main construction as follows:

$$F=\frac{\text{variation between conditions}}{\text{error variance}}=\frac{\text{treatment effects }+\text{ random error}}{\text{random error}}$$

The final F-value for repeated-measures ANOVA, thus, indicates the ratio of variability that is between conditions relative to that which is attributed to random error. The lowest an F-value can be for an ANOVA is 1.00. When F = 1.00, all of the observed variability is attributed to random error and not to the different conditions. This is in keeping with the null hypothesis. However, the greater the F-value, the greater the ratio of variability that is uniquely attributed to the conditions (i.e. the IV) rather than to random differences.

> When there is no treatment effect, the numerator and denominator will be equal and the result will be F = 1.00. Thus, the lowest F-value possible is 1.00 which occurs when no treatment effect is observed. However, the greater the treatment effect, the greater the F-value will be.

When this formula is written out to summarize the computational elements, it is written as follows:

$$F=\frac{MSS_b}{MSS_e}=\frac{SS_{b\div}df_b}{SS_{e\div}df_e}$$

Keep in mind that between calculations for repeated measures ANOVA refer to differences between conditions (i.e. differences between groups of data for the same participants) not differences between different groups of participants. Therefore, in repeated measures ANOVA there can be differences that occur among the participants within each group but there are no differences that occur from having different people in each group.

Sources of Variation in Repeated Measures ANOVA

Let's take a moment to understand the different sources of variation for repeated-measures ANOVA before looking at the way things are calculated so we can understand the rationale for those calculations. There are three things to which variation can be attributed and estimated in ANOVA:

1. **Treatment Effects ($SS_{\text{treatment}}$):** variation attributed to the independent variable (i.e. variation attributed to exposure to different conditions);
2. **Individual Variation (**$SS_{\text{participants}}$ or SS_p): variation attributed to pre-existing, systematic differences among individuals that they bring with them to each condition;
3. **Error Variance (**SS_{error} or SS_e; also known as Residual Error): unexplained variation not accounted for or attributed to either of the aforementioned forms.

It is important to note that the proportions of variance attributed to each of these sources cannot all be directly computed. Instead, some sources can be more directly estimated while others must be deduced by partitioning them out during computations.

Parsing out Sources of Variation

For repeated-measures ANOVA, we must parse out the variation attributed to several sources before the *F*-formula can be used. This can be done using four important computations of variability:

1. $S_{\text{betweenconditions}}$ (SS_b): This includes variation between groups that is attributed to the conditions/treatment (SS_t) and any random error (SS_e) together. This can also be referred to as SS between treatments or simply SS between.
2. SS_{within} (SS_w): This includes variation that exists within group caused by both participant differences that are systematic (SS_p) and random differences (error) which is non-systematic (SS_e).
   - If you compute and sum SS_p and SS_e, you will get SS_w
3. $S_{\text{participants}}$ (SS_p): This includes variation due to differences between participants that they bring with them to each condition.
   - This form of error is part of SS_w but is not part of SS_b.
4. SS_{error} (SS_e): This includes variation that is random and is not attributed to the treatment (i.e. conditions) nor to differences between participants.
   - SS_e exists within both SS_b and SS_w.

SS_b is short for the "sum of squares between conditions." SS_b is the portion of variance attributed to the independent variable (i.e. the different conditions) plus random error (SS_e). SS_w is short for "sum of squares within conditions." There are two sources of variability within conditions that together comprise the $SS_w : 1$. The sum of squares for participants (SS_p) which is the portion of variability attributed to pre-existing differences among people and 2. The sum of squares error (SS_e) which is the variability that is attributed to randomness. Notice that both SS_b and SS_w include random error but only SS_w includes error attributed to differences between people. Therefore, we want to remove the error attributed to pre-existing differences among participants (SS_p) from the SS_w when creating our denominator; this is so that both the numerator and denominator will account for the same sources of error (namely, just the SS_e).

The sum of squares total (SS_T) can also be computed but is not a necessity so we will only briefly review it here. SS_T includes variability from all three sources (i.e. treatment effects, individual variation, and error variance) without partitioning them. SS_T can be computed directly or found by summing SS_b and SS_w because these together contain the variability for all three sources.

Reading Review 11.2

1. What two sources of variance together make up the SS between?
2. What two sources of variance together make up the SS within?
3. What source of variance is included in both the SS between and the SS within?
4. Which SS includes any variance attributed to the independent variable?

11.4: Computations with the Repeated-Measures ANOVA Formula

In order to review how to complete computations for repeated-measures ANOVA we will use Data Set 11.1. Suppose these data were taken from a sample of 8 participants whose confidence was measured under three conditions: before taking a class, immediately after taking a class, and again 6 months after the class. We will use these data to practice testing the hypothesis that mean confidence will be different before (Condition 1), immediately after (Condition 2), and 6 months after taking a class (Condition 3).

Data Set 11.1. Academic Confidence at Three Different Times (n = 8)

Participant ID	Condition 1	Condition 2	Condition 3
A	6	10	8
B	5	8	8
C	5	9	7
D	3	6	6
E	3	6	6
F	4	6	5
G	5	7	6
H	7	10	10

Preparatory Computations

Start by completing descriptive computations as a foundation. When performing computations for each condition, numeric subscripts are used to clarify the condition to which the data and results correspond. For example, the sample size for Condition 1 will be noted as n_1.

First, find the sample size for each condition and the overall size of the data set. For repeated measures ANOVA the sample size refers to the number of raw scores rather than the number of participants. Therefore, the size of the samples for each condition are 8 and the overall size of the data set is 24 (because there are 24 raw scores total across the three conditions). Sample size is indicated with the symbol n and overall size of the data set is indicated with the symbol N. The number or conditions, known as k, should also be noted.

Second, find the means for each condition (e.g. $\bar{X}_1$) and the grand mean ($\bar{X}_{\text{grand}}$). The grand mean is the mean for all raw scores across participants and conditions together.

Third, find the standard deviations for each condition (e.g. s_1). These are not needed when computing an F-value but are useful when reporting results in APA format which is why they should be computed. Note that computing standard deviations within conditions includes computing sums of squares within each condition. Thus, it is most efficient to either reuse the first three steps of your standard deviations computations when computing sums of squares within or to compute standard deviations after computing sums of squares within.

The results of these preparatory computations are shown in the bottom of Table 11.1.

Table 11.1. Preparatory Computations (n = 8)

Participant ID	Condition 1	Condition 2	Condition 3
A	6	10	8
B	5	8	8
C	5	9	7
D	3	6	6
E	3	6	6

Participant ID	Condition 1	Condition 2	Condition 3
F	4	6	5
G	5	7	6
H	7	10	10

Table 11.1. Preparatory Computations (n = 8)

$\bar{x}_{\text{grand}}$ = 6.50	$\bar{x_1}$ = 4.75	$\bar{x_2}$ = 7.75	$\bar{x_3}$ = 7.00
N = 24	n_1 = 8	n_2 = 8	n_3 = 8
k = 3	s_1 = 1.39	s_2 = 1.75	s_3 = 1.60

Sum of Squares Between

Understanding SS_b

Now that the preparatory computations are complete, we can begin our computations for the ANOVA formula, starting with the sum of squares between. In a repeated-measures ANOVA the SS_b is the sum of squares between conditions whereas in a simple, independent groups ANOVA SS_b is the sum of squares between different groups of participants. The formulas and steps used to calculate the SS_b for each of these are identical.

Despite this, however, they include different forms of error due to the differences in how data were collected in their corresponding research designs. Specifically, when the sum of squares between are calculated in repeated-measures ANOVA, they only include two sources of variation: those due to the condition and random variation and do not include error due to using different samples of participants. Because the only source of error contributing to the numerator is random error (i.e. SS_e), only this source of error needs to be addressed in the denominator when using repeated measure ANOVA. We must keep this in mind because we will need to address it when computing the denominator of the F-formula. Now that we have reviewed what SS_b includes for a repeated-measures design, we can proceed to its computations.

The sum of squares between is computed using the same formula for independent-groups ANOVA and repeated-measures ANOVA:

$$SS_b = \Sigma n_i \left[(\bar{x}_i - \bar{x}_{\text{grand}})^2 \right]$$

The resulting value includes treatment effects, random error, and differences between different groups of participants for independent-groups ANOVA.

The resulting value includes only treatment effects and random error for repeated-measures ANOVA.

For this reason, the denominator of the F-formula for independent-samples ANOVA accounts for random error and sample differences using SS_w and the denominator for repeated measures ANOVA only accounts for random error using SS_e.

Computing SS_b

The numerator in the repeated-measures ANOVA is computed using the sum of squares between (SS_b). The formula for computing SS_b is as follows:

$$SS_b = \Sigma n_i \left[(\bar{x}_i - \bar{x}_{\text{grand}})^2 \right]$$

The subscript *i* indicates that the computations will be carried out for each condition (i.e. subgroup of data). Therefore, this formula requires computations for each condition separately which will then be summed to get the overall SS_b. The formula can be translated into simpler language and computed for each condition as follows:

$$SS_{\text{b\_condition}} = n_{\text{condition}} \left[(\bar{x}_{\text{condition}} - \bar{x}_{\text{grand}})^2 \right]$$

The pieces needed are as follows:

$n_{\text{condition}}$: The sample size for the condition

> Note
>
> This should be the same for all conditions because the same sample is being used in each condition

$\bar{x}_{\text{condition}}$: The mean of raw scores for a specific condition such as $\bar{x_1}$

$\bar{x}_{\text{grand}}$: The mean when all data for all conditions are treated as one, grand, group

The sample sizes, condition means, and grand mean can be found in Table 11.1 and can be entered into the formula to find the sum of squares between for each condition. Once the SS_b has been computed for each condition, they are summed together to get the overall SS_b for the test as indicated in the SS_b formula.

> The subscript i stands in for the names of all conditions being tested such that the computations should each be computed for Condition 1, then Condition 2, then Condition 3, and so on until computations for all groups have been completed. Once the computations have been performed for each condition they can be summed to get the overall SS_b for the test.

Computing SS_b is the first major step of using the repeated-measures ANOVA formula and it includes several sub-steps. Let's walk through how to compute SS_b for Data Set 11.1 one sub step at a time:

1. Find the total SS_b across all groups:
   a. Subtract the grand mean ($\bar{x}_{\text{grand}}$) from the mean for Condition 1 ($\bar{x_1}$) to get the deviation for Condition 1.
   b. Square the deviation for Condition 1 from step 1a.
   c. Repeat steps 1a and 1b for each condition to get the squared deviations for each condition.
   d. Sum the squared deviations for all conditions.
   e. Multiply the sum of squared deviation (which is the result of step 1d) by the sample size to find the sum of squares between (SS_b).

These computations for steps 1a through 1e are shown in the bottom of Table 11.2. The final result is SS_b = 39.00.

Table 11.1. Preparatory Computations (n = 8)

Participant ID	Condition 1	Condition 2	Condition 3
A	6	10	8
B	5	8	8
C	5	9	7
D	3	6	6
E	3	6	6
F	4	6	5
G	5	7	6
H	7	10	10
$\bar{x}_{\text{grand}}$ = 6.50	$\bar{x_1}$ = 4.75	$\bar{x_2}$ = 7.75	$\bar{x_3}$ = 7.00
	n_1 = 8	n_2 = 8	n_3 = 8

$$\begin{aligned} SS_b &= 8\left[(4.75-6.50)^2 \quad +(7.75-6.50)^2 \quad +(7.00-6.50)^2\right] \\ SS_b &= 8[3.0625 \quad +1.5625 \quad +0.25] \\ SS_b &= 8[4.875] \\ SS_b &= 39.00 \end{aligned}$$

Sum of Squares Within

Understanding SS_w

When the sum of squares within is computed, it contains the total estimate of pre-existing differences among participants (known as participant variance; SS_p) and unexplained differences (known as error variance; SS_e). However, the repeated-measures ANOVA requires the use of SS_e alone for its denominator (unlike simple ANOVA which uses SS_w for its denominator). However, SS_e cannot be directly computed. Therefore, SS_w must be computed and then partitioned into its two sources of variation in order to find SS_e. To say it another way, we compute the SS_w first so that we can then parse the SS_e from it which is needed for the denominator of the F-formula.

> The denominator for a repeated-measures ANOVA only needs to include SS_e which is part of SS_w. Therefore, SS_w is calculated so that SS_e can then be parsed out from it in a later step.

Computing SS_w

Computing SS_w is the second major step of using the repeated-measures ANOVA formula and it includes several sub-steps. The formula for computing SS_w is as follows:

$$SS_w = \Sigma\left[\Sigma\left(x_i - \bar{x}_{\text{condition}}\right)^2\right]$$

To find SS_w, we must first find the SS_w for each condition and then sum them up. The SS within each condition formula can be written on its own as follows:

$$SS_{w\_\text{condition}} = \Sigma\left(x_i - \bar{x}_{\text{condition}}\right)^2$$

The pieces needed to use the formula are as follows:

x_i: Each raw score in the condition

$\bar{x}_{\text{condition}}$: The mean of raw scores for a specific condition such as $\bar{x_1}$

The raw scores and condition means can be found in Table 11.1 and can be entered into the formula to find the sum of squares within for each condition. $SS_{w\_\text{condition}}$ is found by finding the deviation for each raw score within the condition, squaring those deviations, and then summing them. Once the SS_w has been computed for each condition, they are summed together to get the overall SS_w for the test as indicated in the SS_w formula.

Each condition SS_w can be organized using subscripts (such as SS_w1 for the sum of squares within Condition 1 and so on). Therefore, we must find the SS_w for each condition and then add them all together to find the overall SS_w. If we have three conditions, this can be summarized as follows:

$$SS_w = SS_{w1} + SS_{w2} + SS_{w3}$$

This can be expanded to include as many conditions as needed.

Finding SS_w is the second major step in computing a repeated-measures ANOVA. Let's walk through how to compute SS_w for Data Set 11.1, one sub-step at a time:

2. Find the SS_w for each condition then sum them to get the total SS_w:
   a. Subtract the condition mean ($\bar{x_1}$) from each raw score in Condition 1 to find each deviation.
   b. Square each deviation for Condition 1.
   c. Sum the squared deviations for Condition 1. The result of this step is the SS_w1.
   d. Repeat steps 2a through 2c for each condition until the SS_w is known for all conditions.
   e. Sum the SS_w from all conditions together to get the overall SS_w.

The computations for steps 2a through 2e are shown in Table 11.3. The final result is SS_w = 53.00

Table 11.3. Sum of Squares Within Computations (n = 8)

Participant ID	Condition 1	Condition 2	Condition 3
A	$(6 – 4.75)^2 = 1.5625$	$(10 – 7.75)^2 = 5.0625$	$(8 – 7)^2 = 1$

Participant ID	Condition 1	Condition 2	Condition 3
B	$(5 – 4.75)^2 = 0.0625$	$(8 – 7.75)^2 = 0.0625$	$(8 – 7)^2 = 1$
C	$(5 – 4.75)^2 = 0.0625$	$(9 – 7.75)^2 = 1.5625$	$(7 – 7)^2 = 0$
D	$(3 – 4.75)^2 = 3.0625$	$(6 – 7.75)^2 = 3.0625$	$(6 – 7)^2 = 1$
E	$(3 – 4.75)^2 = 3.0625$	$(6 – 7.75)^2 = 3.0625$	$(6 – 7)^2 = 1$
F	$(4 – 4.75)^2 = 0.5625$	$(6 – 7.75)^2 = 3.0625$	$(5 – 7)^2 = 4$
G	$(5 – 4.75)^2 = 0.0625$	$(7 – 7.75)^2 = 0.5625$	$(6 – 7)^2 = 1$
H	$(7 – 4.75)^2 = 5.0625$	$(10 – 7.75)^2 = 5.0625$	$(10 – 7)^2 = 9$
	$\bar{x_1} = 4.75$	$\bar{x_2} = 7.75$	$\bar{x_3} = 7.00$
	$SS_w1 = 13.50$	$SS_w2 = 21.50$	$SS_w3 = 18.00$
$SS_w = 13.50 + 21.50 + 18.00 + 53.00$			

Sum of Squares Participants

Understanding SS_p

Recall that some of the observed variation within conditions is due to the fact that individual participants have different tendencies. This causes there to be variation in scores within each condition. Let's consider this concept with an example using two participants from Data Set 11.1. Suppose we want to know whether confidence is different under the three different conditions. Suppose that one participant (participant H) tends to have high confidence relative to other participants (such as participant D). Person H has a confidence score that is 4 units higher than person D in all conditions. Here we see the consistency that exists within each person in regards to the dependent variable: One generally has higher scores and the other generally has lower scores. These differences are what are measured as $SS_{\text{participants}}$ (SS_p); these are not caused by the experimental conditions and are not the focus of the hypothesis nor the corresponding ANOVA. Instead, we want to know whether there are differences in the conditions beyond this kind of individual variation in confidence.

Suppose that we find that, on average, both person H and person D were least confident in Condition 1 (before the class), but had higher confidence compared to themselves in both Condition 2 (immediately after the class) and Condition 3 (6 months after the class), even though they had different confidence from each other within each condition. The pattern we are observing here is that there are systematic differences in confidence across (i.e. between) conditions that are not simply due to individual differences; the differences across conditions are attributed to the conditions (via SS_b) and the difference between participants within conditions are a form of error known as SS_p.

When we partition variance in ANOVA, we are isolating amounts of variance attributed to each source so we can focus on the amount which can be attributed to conditions without or relative to other sources (i.e. relative to random error but not including individual differences). When F is computed in a repeated-measures ANOVA, SS_p is computed so that it can be removed from the denominator. This allows us to see the group by group differences (i.e. differences between groups) relative to random differences. Thus, after we compute SS_w, we compute SS_p so that it can be removed.

Computing SS_p

Computing SS_p is the third major step of using the repeated-measures ANOVA formula. The formula for computing SS_p is as follows:

$$SS_p = \Sigma \frac{P^2}{k} - \frac{G^2}{N}$$

We have a few new symbols so let's start by defining each of those and how to compute them.

- P refers to participant totals. This is the sum of raw scores across conditions computed separately for each participant.
- P^2 is found by squaring each participant total separately.

- G refers to the grand total. This is the sum of all raw scores across all conditions for all participants together.
- G^2 is found by squaring the total of all raw scores (i.e. squared G-value).

Let's walk through how to compute SS_p for Data Set 11.1, one sub-step at a time:

3. Find SS_p:
   a. Find each participant total known as P by summing scores across conditions for each participant separately.
   b. Square each P.
   c. Divide each P^2 by *k* (i.e. divide P^2 by the number of conditions).
   d. Sum the results of step 3c
   e. Find G by summing all the raw scores across all conditions.
   f. Square G.
   g. Divide G^2 by N (note: N in repeated-measures ANOVA represents the number of raw scores across all participants and conditions).
   h. Subtract the result of step 3g from the result of step 3d to get SS_p.

The computations for steps 3a through 3h are shown in the bottom of Table 11.4. The final result is SS_p = 48.00.

Table 11.4. Sum of Squares Participants Computations (n = 8)

Participant ID	Condition 1	Condition 2	Condition 3	P (Participant Totals)	P^2	$\frac{P^2}{k}$
A	6	10	8	24	576	576 ÷ 3 = 192
B	5	8	8	21	441	441÷ 3 = 147
C	5	9	7	21	441	441÷ 3 = 147
D	3	6	6	15	225	225÷ 3 = 75
E	3	6	6	15	225	225÷ 3 = 75
F	4	6	5	15	225	225÷ 3 = 75
G	5	7	6	18	324	324÷ 3 = 108
H	7	10	10	27	729	729÷ 3 = 243
Column Totals	38	62	56			$\boldsymbol{\Sigma\frac{P^2}{k} = 1,062}$
G = 38 + 62 + 56 = 156 G^2 = 24,336 This is the total of all raw scores. This is the squared total of all raw scores.						
k = 3 N = 24 This is the number of conditions. This is the number of raw scores across conditions.						

$$SS_p = \Sigma\frac{P^2}{k} - \frac{G^2}{N}$$
$$SS_p = 1,062 - \frac{24,336}{24}$$
$$SS_p = 1,062 - 1,014$$
$$SS_p = 48.00$$

Sum of Squares Error

Understanding and Computing SS_e

SS_e refers to the random error that occurs which is not attributed to differences between participants. It is the focus of the denominator of the F-formula but is not computed directly. Instead, SS_e is one of two sources of error that make up the SS_w and must be parsed from it; SS_w is the total of SS_e and SS_p. Thus, the sum of squared deviations within the conditions (SS_w) minus the sum of squares between participants (SS_p) yields the sum of squares error (SS_e). This represents the otherwise unaccounted for error and is used for the denominator for the repeated measures ANOVA formula. The formula for SS_e is as follows:

$$SS_e = SS_w - SS_p$$

Thus, finding SS_e is simple once both SS_w and SS_p have been calculated. Let's walk through how to compute SS_e for Data Set 11.1:

4. Find SS_e by subtracting SS_p from SS_w. We found SS_p and SS_w in prior steps so we can now simply plug them in and find SS_e as follows:

$$SS_e = 53.00 - 48.00$$
$$SS_e = 5.00$$

The final result is SS_e = 5.00.

Degrees of Freedom

Degrees of Freedom Between (df_b)

df_b is the degrees of freedom between the conditions; it is the adjusted k-value. Recall that k refers to the number of conditions. The formula is as follows:

$$df_b = k - 1$$

df_b is always the number of conditions minus 1. For example, if three conditions were being compared, the df_b would be 2 but if four conditions were being compared the df_b would be 3, and so on. Let's walk through how to compute df_b for Data Set 11.1:

5. Find the df_b by subtracting 1 from the number of groups (k).

$$df_b = k - 1$$
$$df_b = 3 - 1$$
$$df_b = 2$$

Degrees of Freedom Error (df_e)

Degrees of freedom for the error is equal to the adjusted k-value multiplied by the adjusted sample size. Let's walk through how to compute df_e for Data Set 11.1:

6. Find the df_e
   a. Subtract 1 from the number of groups (k)
   b. Subtract 1 from the sample size (i.e. the number of participants)
   c. Multiplying the two resulting values to get df_e

$$df_e = (k - 1)(n - 1)$$
$$df_e = (3 - 1)(8 - 1)$$
$$df_e = (2)(7)$$
$$df_e = 14$$

The final result is df_e = 14.

Putting the Formula Together

Once the four components are calculated, their results are put into the ANOVA formula and used to solve for F.

$$F = \frac{MSS_b}{MSS_e}$$

The numerator asks for the mean sum of squares between conditions (MSS_b). The denominator asks for the mean sum of squares error (MSS_e). Calculating these requires dividing the respective SS by its df, thus, the formula can be rewritten as follows:

$$F = \frac{SS_b \div df_b}{SS_e \div df_e}$$

We computed the four necessary components in prior steps and can now proceed to plugging them in and calculating F. Let's walk through these remaining steps to compute F for Data Set 11.1:

7. Write the ANOVA formula with the four values found in the above steps (i. e. SS_b, df_b, SS_e, and df_e) plugged into their respective locations like so:

$$F = \frac{SS_b \div df_b}{SS_e \div df_e}$$
$$F = \frac{39.00 \div 2}{5.00 \div 14}$$
$$F = \frac{19.50}{0.3571\ldots}$$
$$F = 54.60$$

Repeated Measures ANOVA Computations Summary

The goal of repeated-measures ANOVA is to assess the ratio of treatment effects to that of error variance. Treatment effects (i.e. variability attributed to conditions) are included in SS_b. Error variance is measured as SS_e. These two sources of variance are the focus of the repeated measures ANOVA formula:

$$F = \frac{MSS_b}{MSS_e} = \frac{SS_b \div df_b}{SS_e \div df_e}$$

Numerator Calculations

SS_b is found using the formula:

$$SS_b = \Sigma n_i \left[(\bar{x}_i - \bar{x}_{\text{grand}})^2 \right]$$

1. Use the formula to find the SS_b for each condition.
2. Sum the SS_bs for all conditions to get overall SS_b.

df_b is found using the formula: $k - 1$

Denominator Calculations

The sum of squares error (SS_e) is calculated indirectly by removing the sum of squares attributed to pre-existing differences among participants (SS_p) from the sum of squares within (SS_w) using this formula:

$$SS_e = SS_w - SS_p$$

Therefore, both SS_w and SS_p must be computed before finding SS_e.

SS_w is found using the formula:

$$SS_w = \Sigma \left[\Sigma (x_i - \bar{x}_{\text{condition}})^2 \right]$$

1. Use the formula to find the SS_w for each condition.
2. Sum the SS_ws for all conditions to get overall SS_w.

SS_p is found using the formula:

$$SS_p = \Sigma \frac{P^2}{k} - \frac{G^2}{N}$$

- P refers to participant totals. This is the sum of raw scores across conditions computed separately for each participant.
- G refers to the grand total. This is the sum of all raw scores across all conditions for all participants together.

Then SS_e is found using the formula:

$$SS_e = SS_w - SS_p$$

Next, df_e should be calculated using the formula: $(N-k)(n-1)$

F Calculations

MSS_b is found using the formula:

$$MSS_b = SS_b \div df_b$$

MSS_e is found using the formula:

$$MSS_e = SS_e \div df_e$$

F is found using the formula:

$$F = \frac{MSS_b}{MSS_e}$$

Reading Review 11.3

1. What is a grand mean and how is it calculated?
2. What is P in a repeated-measures ANOVA and how is it calculated?
3. What two sources of variance together make up the SS between for repeated-measures ANOVA?
4. What two sources of variance together make up the SS within for repeated-measures ANOVA?
5. What is being calculated and represented by the numerator of the repeated-measures ANOVA formula?
6. What is being calculated and represented by the denominator of the repeated-measures ANOVA formula?

11.5: Testing a Hypothesis with Repeated Measures ANOVA

The computations have been reviewed in detail in the prior section. In addition, this chapter is focused on using the F-formula to test a non-directional hypothesis. Therefore, this section will briefly review the most relevant steps for testing a non-directional hypothesis and will refer to the prior section of this chapter for the steps to using the F-formula.

Steps in Hypothesis Testing

In order to test a hypothesis, we must follow these steps:

1. **State the hypothesis.**

The hypothesis for Data Set 11.1 can be declared as follows: *It is hypothesized that the mean confidence of a group of students will be different under three different conditions: before a class, immediately after the class, and 6 months after the class.*

A one-tailed test is used when conducting the omnibus test in ANOVA, regardless of whether the hypothesis is stated directionally.

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The means of one sample measured under three conditions are being compared so the appropriate test is a repeated-measures ANOVA.

3. **Determine the critical value.**

In order to determine the critical value for a repeated-measures ANOVA, three things must be identified:

a. the alpha level,
b. the Degrees of Freedom Between (df_b), and,
c. the Degrees of Freedom Error (df_e).

The default alpha level of .05 is appropriate because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The df_b and the df_e are reported here (see the prior section for computations):

$$df_b = 2$$
$$df_e = 14$$

These three pieces of information are used to locate the critical value from the test. The full tables of the critical values for F-tests are located in Appendix E. Under the conditions of an alpha level of .05, df_b = 2, and df_e = 14, the critical value is 3.739.

$$CV = 3.739$$

The critical value represents the value which must be exceeded in order to declare a result significant. Thus, in order for the result to significantly support the hypothesis is needs to exceed the critical value of 3.739.

4. **Calculate the test statistic.**

A test statistic can also be referred to as an obtained value. The formula needed to find the test statistics (F) for this scenario is as follows:

$$F = \frac{MSS_b}{MSS_e}$$

Computations are detailed in in the prior section of this chapter and only the result is shown here. The obtained value for this test was 54.60 when rounded to the hundredths place.

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value for F exceeds the critical value as follows:

The critical value is 4.737.

The obtained F value is 54.60.

The obtained F-value exceeds (i.e. is greater than) the critical value. Therefore, the criteria has been met to declare the result significant. Thus, the result significantly supports the hypothesis.

> **Note**
>
> If the hypothesis had a direction, the directions of each group by group comparison would need to be checked in post-hoc analyses before concluding that all aspects of the hypothesis had been supported. However, the current hypothesis was simply that groups would differ. Therefore, we are able to conclude that the hypothesis was supported without needing post-hoc results. However, it is still good practice to report the means and standard deviations for each condition in our APA-formatted summary of the results.

6. **Calculate the effect size and/or other relevant secondary analyses.**

When it is determined that the result is significant, effect sizes and post-hoc analyses may be computed. We reviewed effect size and post-hocs for ANOVA in Chapter 10. Here we will briefly note changes that needs to be made to the formulas for effect size and post-hoc formulas when using repeated-measures ANOVA. However, these are not a focus of this chapter and the post-hocs will not be computed for the present hypothesis because it was non-directional. Instead, the adjustments that must be made to these computations when using repeated-measures ANOVA are included here for your reference.

The effect size is a calculation of the percent of variance observed that was systematic (and, thus, was uniquely between groups) which is reported in decimal form (see Chapter 10 for a review of this computation). The symbol for this effect size is η^2(which is named "eta squared"). An amended version of the formula is used for repeated measures ANOVA which removed participant differences. This is because participant differences (SS_p) are not captured in sum of squares between when using a repeated measures ANOVA, and, therefore, need to be removed from the denominator of the formula. Therefore, effect sizes for repeated-measures ANOVAs are computed with the amended formula as follows:

$$\eta^2 = \frac{SS_b}{SS_T - SS_p}$$

For the same reason, the denominator for the post-hoc formula is also changed when it accompanies a repeated-measures ANOVA. Specifically, the mean sum of squares error (MS_e) is used for the numerator of the HSD formula for repeated-measures ANOVA instead of the mean sum of squares within (MS_w) which was used for the simple, independent groups ANOVA (see Chapter 10 for review of post-hoc analyses). Therefore, the effect size *HSD* formula for repeated-measures ANOVA is as follows:

$$HSD = q\sqrt{\frac{MS_e}{n}}$$

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

Keep in mind that results are reported in past tense because they report on what has already been found. In addition, the research hypothesis must be stated but the null hypothesis is usually not needed for summary paragraphs because it can be deduced

from the research hypothesis. Finally, APA format requires a specific format be used for reporting the results of a test. This includes a specific format for reporting relevant symbols and details for the formula and data used. Following this, the results for our hypothesis with Data Set 11.1 can be written as shown in the summary example below.

> **APA Formatted Summary Example**
>
> A repeated-measures ANOVA was used to test the hypothesis that the mean confidence of a group of students would be different under three different conditions: before a class, immediately after the class, and 6 months after the class. The mean confidence scores were significantly different before (M = 4.75, SD = 1.39), immediately after (M = 7.75, SD = 1.75), and 6 months after a class (M = 7.00, SD = 1.60), $F(2, 14) = 54.60, p < .05$. Thus, the hypothesis was supported.

Reading Review 11.4

1. What information is needed to find the critical value for a repeated-measures ANOVA?
2. How is significance determined for a repeated-measures ANOVA?

11.6: Structured Summary for the Repeated-measures ANOVA

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes. Subscript values are used to note conditions assuming three conditions were used. Expand as needed when more than three conditions are used.

Summarize what each symbol stands for, assuming three groups are being compared.

n_1 =

n_2 =

n_3 =

N =

k =

$\bar{X}_1$ =

$\bar{X}_2$ =

$\bar{X}_3$ =

$\bar{X}_{\text{grand}}$ =

s_1 =

s_2 =

s_3 =

SS_{b1} =

SS_{b2} =

SS_{b3} =

SS_b =

> **Note**
>
> SS_b in repeated measures ANOVA refers to differences between conditions due to the treatment and those due to random error (SS_e). Unlike in an independent groups ANOVA, there is no additional variation in SS_b caused by having different participants in each group because the same participants are used in all groups in a repeated measures design. Thus, the computations for SS_b are the same in each version of ANOVA, but it is important to keep in mind that they do not both include error caused by using independent groups in conditions.

SS_{w1} =

SS_{w2} =

SS_{w3} =

SS_w =

> **Note**
>
> SS_w in repeated measures ANOVA includes pre-existing participant differences (SS_p) and random error (SS_e).

P =

G =

SS_p =

SS_e =

df_b =

df_e =

MSS_b =

MSS_e =

Note

MSS_e is used as the denominator in repeated measures ANOVA rather than MSS_w which is used in independent groups ANOVA.

Fill-in the appropriate information for each section below:

1. Repeated-Measures ANOVA Basics
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What assumptions must the data meet to use this test?
2. Repeated-Measures ANOVA Formula
   a. What is the formula for a repeated-measures ANOVA?
   b. What things should be computed in the preparatory steps?
   c. What are the steps for solving using this formula?
3. Reporting Results from a Repeated-Measures ANOVA
   a. How is this statistic reported when using APA format?
      i. What things must be reported in the APA summary sentence for the F test?
      ii. What two specific statistics should be reported for each of the conditions?

CHAPTER OVERVIEW

12: Bivariate Correlation

12.1: Introduction to Bivariate Correlation

The **bivariate correlation** is used when you want to test whether two quantitative variables are related. Related in this sense refers to there being a linear pattern between the two variables. This is conceptually distinct from what we have seen so far in this book. In our prior chapters (Chapters 7 through 11) we focused on comparing something quantitative between groups or conditions. However, not all hypotheses and data sets have a qualitative grouping or condition variable. Instead, there are times when the data are only quantitative and we wish to analyze those variables together. When this occurs, bivariate correlation may be the best fit to the hypothesis and data. For example, a bivariate correlation could be used to test whether income is related to level of happiness or whether hours spent exercising is related to amount of stress. In each of these examples the relationship between two quantitative variables (income and happiness in the first example and exercise and stress in the second example) is the being proposed and, thus, would be tested.

The version of correlation which is the focus of this chapter is known as a Pearson's Product Moment Correlation (PPMC) or a Pearson's Correlation. Because it is so commonly used, it is often simply referred to simply as a correlation without specifying the full name.

12.2: Variables

The bivariate correlation is best used to test the relationship between two quantitative variables. Though there are some other ways correlation can be used, this standard bivariate correlation will be the focus of this chapter. Each quantitative variable should be measured using an interval or ratio scale of measurement. The two variables are often referred to as X and Y when spoken about generally. The scores for these variables are referred to as X-values and Y-values, respectively.

12.3: Data and Assumptions

Each statistical test has some assumptions which must be met in order for the formula to function properly. In keeping, there are a few assumptions about the data which must be met before a bivariate correlation is used. First, the scores for both variables must be matched. This means that participants have scores for each variable which are identified together. Second, data for each quantitative variable should be fairly normally distributed without notable impact due to outliers, including bivariate outliers. A bivariate outlier refers to a participant whose scores on X and Y together do not follow the pattern of the other participants. In order to be identified as an outlier, this divergence should be fairly dramatic and rare relative to the rest of the data set rather than mildly divergent or common. Finally, the relationship between the variables should be linear meaning they approximate a line rather than other shapes such as what would be seen when graphing a quadratic equation. We will focus on how to look for the shape visually to see if it looks linear later in this chapter.

12.4: Hypotheses

Unlike *t*-tests and ANOVAs, correlation is not used to assess differences and similarities among different groups or conditions. Instead, it is used to assess whether two quantitative variables are related and, if so, in which way and how strongly. Hypotheses for correlation can be non directional, in which case they require a two-tailed test of the hypothesis, or directional, in which case they require a one-tailed test of the hypothesis.

For the bivariate correlation, the non-directional research hypothesis is that the two variables will be related. The corresponding null hypothesis is that the two variables are not related. The relationship between variables is summarized with the symbol r which can range from 0 (indicating no relationship) up to positive or negative 1.00 (each of which indicate a perfect relationship). However, hypotheses are about truths beyond the sample so population symbols should be used for those. The symbol for correlation with samples is *r* and for populations is the Greek letter *ρ* (named “rho”). Thus, ρ will be use in place of r when hypotheses are written in symbol format. The non-directional research and corresponding null hypotheses for bivariate correlation can be summarized as follows:

Non-Directional Hypothesis for a Bivariate Correlation

Research hypothesis	Variable X will be related to Variable Y.	$H_A: \rho_{xy} \neq 0$
Null hypothesis	Variable X will not be related to Variable Y.	$H_0: \rho_{xy} = 0$

There are two directional hypotheses possible for bivariate correlation. The first is that there will be a positive relationship between the two variables. This version of the research and corresponding null hypotheses can be summarized as follows:

Positive, Directional Hypothesis for a Bivariate Correlation

Research hypothesis	Variable X will be positively related to Variable Y.	$H_A: \rho_{xy} > 0$
Null hypothesis	Variable X will not be positively related to Variable Y.	$H_0: \rho_{xy} \leq 0$

The second possible directional hypothesis is that there will be a negative relationship between the two variables. This version of the research and corresponding null hypotheses can be summarized as follows:

Negative, Directional Hypothesis for Bivariate Correlation

Research hypothesis	Variable X will be negatively related to Variable Y.	$H_A: \rho_{xy} < 0$
Null hypothesis	Variable X will not be negatively related to Variable Y. ***Note:*** *this could also be worded as “Variable X will be positively related or unrelated to Variable Y.”*	$H_0: \rho_{xy} \geq 0$

12.5: Interpretation of r-Values

The symbol for the results of a correlation test is r. The result is known as an r-value, a *correlation coefficient*, or simply as the obtained value for the test. This is easy to remember because r is used to summarize the *relationship* between two variables. The names of variables, or placeholders for their names, are often shown is subscript. Thus, r_{XY} simply indicates that a correlation is being computed between one variable, referred to as X, and another, which is referred to as Y. We can see this reflected in the symbol-formatted hypotheses in the prior section. If the names of the variables are known, those can be used instead of these generic placeholders. For example, $r_{\text{sleep\_stress}}$ could be used when a correlation is being computed between hours of sleep and level of stress. An underscore or hyphen can be used to distinguish between the two variable names in the subscript for ease of comprehension.

The r-value provides two important summaries about the relationship between two variables. It indicates:

1. The direction of the relationship and
2. The strength of the relationship.

Let's take a moment to review each of these and how they are interpreted from the r-value.

Direction of a Correlation

Correlations summarize linear relationships. When a relationship exists between two variables, that relationship can be either positive or negative.

Positive correlations refer to patterns where relatively high values of one variable tend to co-occur with relatively high values of the other variable and relatively low values of one variable tend to co-occur with relatively low values of the other variable. Another way to say this is that scores for two variables have a tendency to increase or decrease together. Positive correlations can also be called direct correlations.

Negative correlations refer to patterns where relatively high values of one variable tend to co-occur with relatively low values of the other variable and relatively low values of one variable tend to co-occur with relatively high values of the other variable. Another way to say this is that scores for one variables have a tendency to decrease when scores from the other variable increase. Negative correlations can also be called indirect correlations or inverse correlations.

Patterns Indicated by Correlation Coefficients

Positive Correlations	Negative Correlations
XY and XY	XY and XY

Patterns Indicated by Correlation Coefficients

↑↑ and ↓↓	↑↓ and ↓↑

Let's consider a few examples to put these two possible directions into context. Consider the possible correlation that could exist between sleep and quiz scores. The sample is comprised of college students and the two variables being measured are hours of sleep and quiz scores. If there is a positive correlation between these two variables, it would mean that as sleep increased across the sample, quiz scores also tended to increase; thus, when this is the case higher scores tend to co-occur with higher scores and lower scores tend to co-occur with lower scores. When the two variables are trending in the same way, their relationship is positive.

In contrast, consider a possible correlation that could exist between caffeine consumption and sleepiness among adults. The sample would be comprised of adults and the two variables being measured would be caffeine and sleepiness, each measured quantitatively. If there is a negative correlation between these two variables, it would mean that as caffeine intake increased across the sample, sleepiness tended to decrease. In this example, higher scores tend to co-occur with lower scores and lower scores tend to co-occur with higher scores. Because they are trending in opposition to each other (in opposite directions), their relationship would be negative.

Direction Using Graphs

Data for correlations can be graphed using a scatterplot (also known as a scattergram or scatter dot graph). Data from the first quantitative variable are plotted using the x-axis and data from the second quantitative variable are plotted using the y-axis. The x-axis runs horizontally and the y-axis runs vertically. Thus, the two scores for each case are used together as a coordinate pair to place a dot on the two-dimensional graph. Coordinate pairs are written generally as (x, y) indicating that the first value in each pair specifies location on the x-axis and the second specifies location on the y-axis.

If there is a relationship between the variables, the data should approximate the shape of a line either angling up or down from left to right. The angle of a linear graph is known as its slope. In correlation, the slope is used to determine whether a relationship is positive or negative. If the graph angles up from left to right, the slope and the corresponding correlation will be positive. In contrast, if the graph angles down from left to right, the slope and corresponding correlation will be negative.

Let's take a look using an example. Suppose a researcher collected data from 10 college students about their hours of sleep and scores on a quiz to test the hypothesis that sleep would positively relate to quiz scores. The scores are shown in Data Set 12.1 below. The first column is not a test variable. Instead, it shows anonymous names (or identification numbers) for each case in the sample. These are sometimes used in research to help connect scores across variables for participants. Each row represents a different participant (i.e. case). The second column shows the data for sleep hours and the third column shows the data for quiz scores.

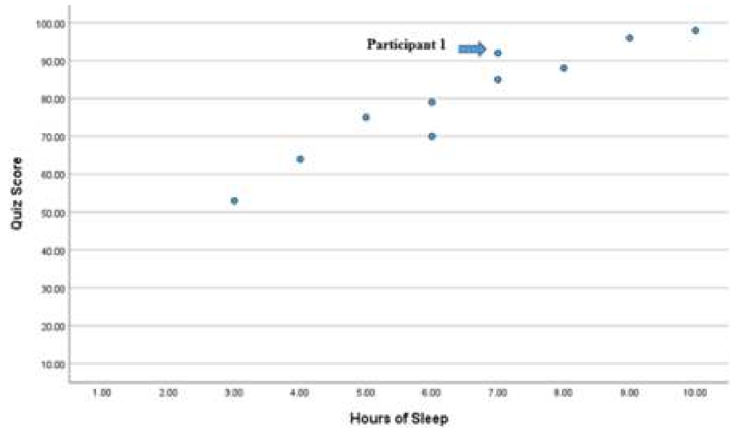
Data Set 12.1. Hours of Sleep and Quiz Scores (n = 10)

Participant Number	Sleep Hours	Quiz Score
1	7	92
2	8	88
3	9	96
4	6	70
5	6	79
6	4	64
7	5	75
8	10	98
9	3	53
10	7	85

Before we graph the data, we can use a simple method to get a sense of the possible direction of the correlation. To do so, we identify the higher and lower scores in each variable to see what the pattern is between the variables, if any. The five highest hours of sleep are 10, 9, 8, 7 and 7 and the lowest are 3, 4, 5, 6 and 6. The highest five quiz scores are 98, 96, 92, 88 and 85 and the lowest are 53, 64, 70, 75, and 79. In Data Set 12.1, the five highest scores for the X variable occurred in the same cases as the five highest scores for the Y-variable. Consistent with this, the lowest five scores for the X-variable occurred in the same cases as the five lowest scores for the Y-variable. Thus, a positive pattern is apparent between the variables so we should expect to see a positive slope when graphed.

We can also see this pattern by graphing. Below is a scatterplot for Data Set 12.1. Each dot represents a case or participant. For example Participant 1 had an x-value of 7 and a y-value of 92 giving them a coordinate pair of (7, 92). You can see their location marked on the graph as an example. The location is over 7 along the x-axis and up 92 along the y-axis. The same process can be used to identify each participant on the graph.

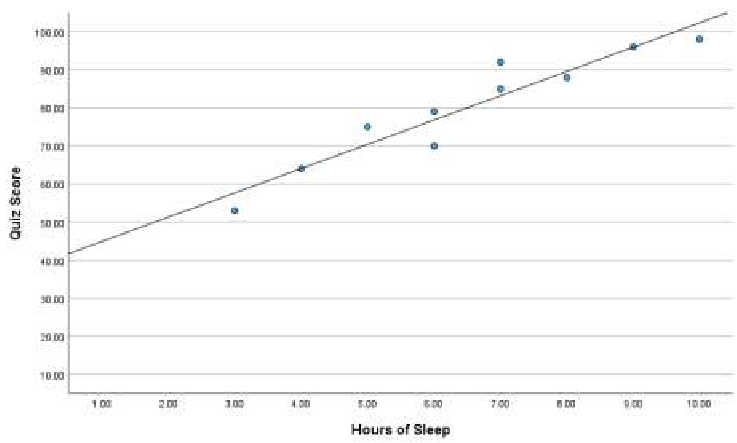
If two variables are correlated, we should see the lines approximating a straight line either sloping up or down from left to right. When we look at the pattern of the dots (which represent the data) in Graph 12.1, we can see that they are approximately following the path of a straight line angling up.



Graph 12.1. Hours of Sleep and Quiz Scores (*n* = 10)

To know whether the correlation is positive or negative we can visually assess the slope. We can see that the line is angling up from left to right, which means that the slope is positive. This is because slope refers to how much the graph rises along the y-axis relative to how much it increases as it runs (i.e. moves) across the x-axis. This is often summarized by saying "slope equals rise over run." Because we read the x-axis from left to right, the run is increasing (meaning it has a positive value). When we read the graph from left to right, the slope is the change in the line relative to the y-axis. If the dots are going up from left to right, the relationship is positive. If the dots were going down from left to right, the relationship would be negative.

To make the slope clearer, a fit line is often added to a graph. A fit line can also be called a line of best fit or slope line in correlation (in regression this same line is called a regression line). A fit line is balanced to approximate the location of the dots with as little error as possible. Below is the scatterplot for Data Set 12.1 with a fit line added. Notice that we can clearly see that the line is angling up (or rising) from left to right and that the correlation is, therefore, positive.



Graph 12.2. Hours of Sleep and Quiz Scores with Fit Line

Magnitude of a Correlation

Correlations have a magnitude, also known as a strength. Thus, magnitude indicates how strongly the two variables are related to each other. The clearer and more consistent the pattern between the variables, the stronger their correlation. The strength is estimated using *r* values and can be understood visually using scatterplots.

Magnitude Using r

The strength of a correlation is summarized using the number of an *r*-value without consideration for the sign of the value (because the sign is the direction and the value is the strength). All correlation strengths are between .00 and 1.00. They cannot exceed these boundaries, in absolute value. When there is no correlation between two variables, the correlation strength is .00. When there is a moderate correlation between two variables, the correlation strength will be at or around .50. When there is a perfect correlation between two variables, which is the strongest a correlation can be, the correlation strength will be 1.00. Thus, the closer r is to .00, the weaker it is and the closer r is to 1.00, the stronger it is.

It can be useful to interpret and describe the strength of a correlation. Though there are no definite rules on how the strength of a specific r-value must be described, there are general guidelines that can be used. These provide approximate ranges and how their strengths can be described, but we must keep in mind that these are guidance and that the most appropriate wording can depend upon the implications of the correlation or other considerations in research and theory relative to the variables. Thus, the cutoffs and terms used here should be considered useful but not obligatory or absolute in their meaning.

Table 12.1. Guidelines for Interpreting the Strength of an r-Value

r-value	Description of Strength
1.00	Perfect
~ .80 to.99	Very Strong
~ .60 to .79	Strong
~ .40 to .59	Moderate
~ .20 to .39	Weak
~ . 00 to .19	No Correlation to Very Weak

Remember, strength is determined by the magnitude of the result, regardless of its direction. Here are some examples of r-values, their magnitudes, and the description of their strengths following the guidelines:

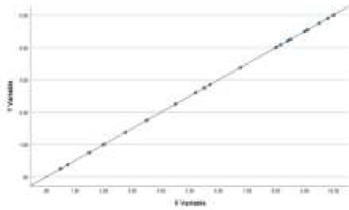
r-value	Magnitude	Description of Strength
.93	.93	Very Strong
-.27	-.27	Weak
-1.00	1.00	Perfect
.46	.46	Moderate
.11	.11	Very Weak
-.11	.11	Very Weak
.68	.68	Strong
.00	.00	No Correlation

Magnitude Using Graphs

The magnitude can also be roughly estimated by looking at the nature of the graph, though this does take some practice. However, reviewing graphs can help us get a sense of the pattern between the variables and a better understanding of what a correlation is computing. The values used for the x-axis, the y-axis, and the sample size can all impact the way a scatterplot looks, making it hard to compare graphs which vary on any or all of these three things. Using different directions of the slope (positive compared to negative) can also make comparisons a bit more challenging. Therefore, we will review several graphs with different magnitudes while keeping the axes, direction (positive), and sample sizes consistent to allow us to more easily focus on the magnitudes.

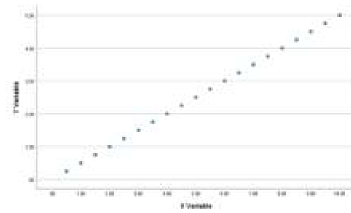
We will look at graphs starting from a perfect correlation and gradually move to no correlation so you can observe how magnitude is reflected visually. Each dot represents a case (or participant). The more closely the dots approximate a straight line, the stronger the relationship is. This means that when a fit line is used, dots are closer to the line when the correlation is stronger. When the dots are more spread out such that they do not tend toward forming a line, the correlation is weaker. This means that when a fit line is used, dots are further from the line when the correlation is weaker.

Thus, strength can be assessed visually by looking at how close the dots are to forming a perfectly straight line that angles up or down from left to right. Take a look at Graphs 12.3 and 12.4 below. Both of these have a perfect correlation. In Graph 12.3, we see a perfectly uniform line angling up from left to right. Even without a fit line added, the straight line is clear. A fit line has been added to Graph 12.4 to illustrate that it is also perfectly straight, despite the data points being unevenly spaced along the line. Both of these are perfect, positive correlations.



Graph 12.3. Perfect, Positive Correlation

(r = 1.00)

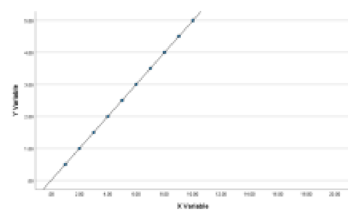


Graph 12.4. Perfect, Positive Correlation

(r = 1.00)

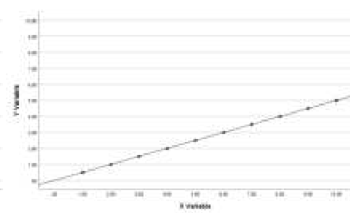
It is important to note that a common mistake people make is assuming the steepness of the line represents the magnitude; however, steepness can be a bit misleading. This is because the steepness of the line will look different dependent upon the anchors used to label the x-axis and y-axis. Thus, it is possible to make the same data look steeper or flatter by changing the height of the y-axis or the width of the x-axis to compress or stretch the look of the graph. Therefore, we will focus on how close the dots are to the line rather than how steep or flat the line looks when visually assessing magnitude.

Let's take a moment to look at how the steepness can be misleading. The two examples below show how Graph 12.3 would look if the x-axis anchors were changed and if the y-axis anchors were changed. Notice that the line seems steeper in one and less steep in the other but that the locations represented by each dot and how straight and perfect the lines are have not changed. All three versions of Graphs 12.3 were made using the same data set and all have the same correlation coefficient of r = 1.00. Thus, the steepness of the slope alone is not a great indicator of strength. Instead, strength is assessed visually by looking at how closely the dots approximate a line and checking that the line is not completely horizontal. As long as a line is not completely flat (meaning as long as it is not completely parallel to the x-axis), it is possible for the two variables to be related.



Graph 12.3 with Altered X-Axis

(r = 1.00)

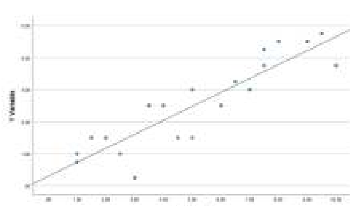


Graph 12.3 with Altered Y-Axis

(r = 1.00)

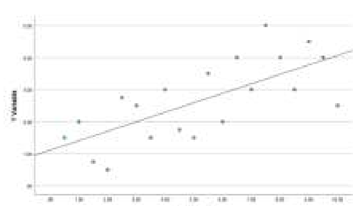
To get a sense of how to estimate the strength of a correlation based on a graph, we will look at several scatterplots with the same sample size but with varying magnitudes. Because the visual steepness of the line can be impacted by changes to the anchors on the axes (as we just saw), the same values are used for both the x- and y-axes for each comparison graph.

Very strong correlations have dots that are relatively close to the line, but without all dots falling in a perfectly straight line. Strong correlations have dots that still clearly trend in a line but with noticeably more distance between the dots and the line, on average. Note that it is the average distance between the dots and the line that matters. In keeping, notice that, overall, the dots are closer to the line and that the correlation coefficient (r-value) is stronger (greater) in Graph 12.5 than in Graph 12.6.



Graph 12.6. Very Strong, Positive Correlation

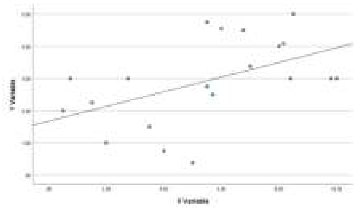
(r = .90)



Graph 12.6. Strong, Positive Correlation

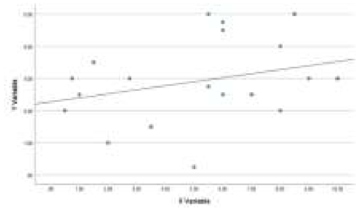
(r = .70)

As the data, on average, are farther from approximating a straight line, the correlation is weaker. Notice that as we move in order from Graphs 12.5 to 12.8, the data are spread farther and farther from the fit line. The less the data approximate a straight line angling either consistently up or consistently down, the weaker the correlation is.



Graph 12.7. Moderate, Positive Correlation

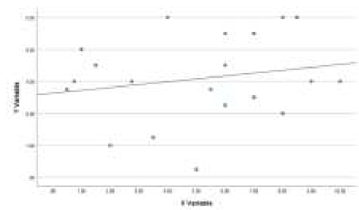
(r = .50)



Graph 12.8. Weak, Positive Correlation

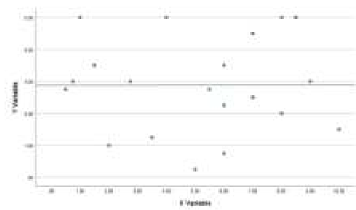
(r = .30)

By the time we get to Graph 12.9, it would be very hard to even see whether the data are generally following a line without the fit line there to guide the eye, because the data are only very loosely following a positive sloping pattern. Graph 12.9 has a very weak correlation. In Graph 12.10 there is no relationship present. The data appear scattered about with no clear slope which they tended to follow. Thus, when there is no relationship between the variables, the fit line lies perfectly horizontal, parallel with the x-axis.



Graph 12.9. Very Weak, Positive Correlation

(r = .10)



Graph 12.10. No Correlation

(r = .00)

Interpreting r

Graphs provide us the ability to see the correlation visually whereas correlation coefficients (r values) summarize both the direction and strength of those correlations. These are two ways to represent the relationships between two quantitative variables. However, the brilliant construction of the correlation coefficient formula makes interpretations easy and allows for the easy comparison of different correlations. Thus, correlations are usually interpreted simply by looking at their r-values.

Let's take a look at some examples. Correlations of .75 and -.75 have the same magnitude but different directions. When r =.75, the correlation is strong and positive. When r = -.75, the correlation is strong and negative. A correlation of r = -.93 is very strong and negative. A correlation of r =.28 is weak and positive. A correlation of r =.06 is very weak and positive. We can imagine the approximate look of the graph, including whether it slopes up or down and how close the dots are to forming a line, simply by knowing the r-value.

Limitations

A major limitation of the correlation is that it cannot be used to determine cause-effect relationships. Just because two things are mathematically related, does not mean that either is the cause of the other. Therefore, though tempting, it is usually inappropriate to use causal language when interpreting the results of correlation (see Chapter 8 for a review of causal language).

Reading Review 12.1

1. What assumptions must be met before using a Pearson's Product Moment Correlation (PPMC)?
2. What is a non-directional hypothesis that could be tested using a Pearson's Product Moment Correlation (PPMC)?
3. What is the symbol used for correlation coefficients?
4. How would the strength and direction of each of the following correlation coefficients be described, using the general rules of thumb for strengths: -1.00, -.44, .59, .12, .86, -.70, .70

12.6: The Bivariate Correlation Formula

The correlation coefficient is used to summarize the relationship between two quantitative variables in a dataset using a number ranging from -1.00 to 1.00. Recall that variables, by definition, vary. Covariance, therefore, refers to the extent to which two variables vary together in a patterned way. When a correlation is perfect, it means that X and Y have a perfect pattern of covariance such that the amount of covariance is equal to the total of the individual variability of the two variables. Thus, the correlation formula is assessing the proportion of shared variance to unshared variance. When r is stronger, it means a greater proportion of the total variance is shared. When the correlation coefficient is 1.00 it means that all of the variance is shared between the variables. When the correlation coefficient is .00 it means that all the variance is attributed to the two variables separately and is, thus, not indicative of a pattern or relationship between the variables. To summarize, we can understand the formula's main construction and outcomes as follows:

$$r=\frac{\text{how greatly } X \text{ and } Y \text{ vary together}}{\text{how greatly } X \text{ and } Y \text{ vary separately}}=\text{ratio of covariance to random variance}$$

When this formula is written out to summarize the computational elements, it is written as follows:

$$r=\frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2}}$$

The numerator of the formula focuses on deviations of each variable (which are a central part of variance) and their connections. The denominator of the formula looks at the deviations for each variable separately. When put together, the result indicates how much of the variance was shared verses separate. The greater the magnitude of final result, the stronger the covariance and, thus, the stronger the relationship is between the two variables.

Notice that the sum of squares within formulas have, once again, made their way into an inferential formula. You can see $\Sigma(X-\bar{X})^2$ in the left side of the denominator which is the sum of squared deviations for the X values. You can also see $\Sigma(Y-\bar{Y})^2$ in the right side of the denominator which is the sum of squared deviations for the Y values. Thus, the correlation formula can also be written as follows by replacing those two SS formulas with the symbol for their respective SSs, as follows:

$$r=\frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{SS_X}\sqrt{SS_Y}}$$

Formula Components

Now that we have taken some time to understand the basic construction of the correlation formula, let's focus on how to actually use it, starting with identifying all of its parts.

In order to solve for r, we need three things:

$\Sigma(X-\bar{X})(Y-\bar{Y})$ = the sum of the products of the deviations from X and Y

$\Sigma(X-\bar{X})^2$ = sum of squared deviations for the X-values

$\Sigma(Y-\bar{Y})^2$ = sum of squared deviations for the Y-values

Formula Steps

The steps are shown in order and categorized into two sections:

A. preparation and
B. solving.

Preparation steps for correlation include finding the mean for X and the mean for Y. Then, these are used in section B to find deviations needed for the three main formula components listed above. Once those components are known and plugged into the formula, order of operations is followed to yield the obtained value (known as r) for the formula. Follow these steps, in order, to find r:

Section A: Preparation

1. Find $\bar{X}$ (the mean for the X-variable scores)
2. Find $\bar{Y}$ (the mean for the Y-variable scores)

> **Note**
>
> Though we do not need to find n for this version of the formula, it is good to make note of what it is because it will be used to find the degrees of freedom (df) later.

Section B: Solving

The values from the preparatory steps must now be used to find the three main components of the correlation formula before the r-value can be computed.

1. Find $\sqrt{\Sigma(X-\bar{X})^2}$
   a. Find $(X-\bar{X})$ by subtracting the mean of X from each x-value (these are the deviations for each X)
   b. Find $\Sigma(X-\bar{X})^2$ by squaring each deviation and then summing those values (which yields the sum of squares for the X-variable, also known as SS_X)
   c. Square root the SS_X.
2. Find \(\sqrt{\Sigma(Y-\bar{Y})^2\)
   a. Find $(Y-\bar{Y})$ by subtracting the mean of Y from each y-value (these are the deviations for each Y)
   b. Find $\Sigma(Y-\bar{Y})^2$ by squaring each deviation and then summing those values (which yields the sum of squares for the Y-variable, also known as SS_Y)
   c. Square root the SS_Y.
3. Find $\Sigma(X-\bar{X})(Y-\bar{Y})$
   a. Find $(X-\bar{X})(Y-\bar{Y})$ by multiplying each deviation from X by its deviation from Y to get the product of the deviations for each case
   b. Find $\Sigma(X-\bar{X})(Y-\bar{Y})$ by summing the product of deviations for each case (i.e. sum the results from step 3a). This is the numerator for the formula.
4. Find the denominator by multiplying the square root of SS_X (which is the result of step 1c) by the square root of SS_Y (which is the result of step 2c)
5. Divide the numerator (which is the result of step 3b) by the denominator (which is the result of step 4).

Reading Review 12.2

1. What is the focus of the numerator of the correlation formula?
2. What is the focus of the denominator of the correlation formula?
3. Which two descriptive statistics should be found in preparation for using the correlation formula?
4. What are the steps to calculating SS_X?
5. What are the steps to calculating SS_Y?

Example of How to Test a Hypothesis Using Correlation

Let's test the hypothesis and Data Set 12.1, which was introduced earlier in this chapter. We supposed a researcher collected data from 10 college students to test the hypothesis that sleep would positively relate to quiz scores. Assume that Data Set 12.1 includes data from the aforementioned sample. Let's follow the steps in hypothesis testing using these data.

Data Set 12.1. Hours of Sleep and Quiz Scores (n = 10)

Participant Number	Sleep Hours	Quiz Score
1	7	92
2	8	88
3	9	96
4	6	70

Participant Number	Sleep Hours	Quiz Score
5	6	79
6	4	64
7	5	75
8	10	98
9	3	53
10	7	85

Steps in Hypothesis Testing

In order to test a hypothesis, we must follow these steps:

1. **State the hypothesis.**

A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that hours of sleep will positively relate to quiz scores.* The format shown in the table below could also be used. Because the hypothesis is directional a one-tailed is needed.

Directional Hypothesis for a Bivariate Correlation

Research hypothesis	Hours of sleep will be positively related to quiz scores.	$H_A : r_{xy} > 0$
Null hypothesis	Hours of sleep will not be positively related to quiz scores.	$H_0 : r_{xy} \leq 0$

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The relationship between two quantitative variables is being tested so the appropriate test is a bivariate correlation.

3. **Determine the critical value.**

In order to determine the critical value for a bivariate correlation, three things must be identified:

1. the alpha level,
2. the degrees of freedom (df), and
3. whether the hypothesis is directional (requiring a one-tailed test) or non-directional (requiring a two tailed test).

The alpha level is often set at .05 unless there is reason to adjust it such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The df must also be calculated. The df is calculated as the sample size minus the number of variables being tested. In bivariate correlation, there are always 2 variables being tested so the formula is $df = n – 2$. The sample size in Data Set 12.1 is 10. Thus, the df for Data Set 12.1 is as follows:

$$df = n - 2$$
$$df = 10 - 2$$
$$df = 8$$

The hypothesis is directional because it specified that the expected correlation would be positive. Thus, this hypothesis requires a one-tailed test of significance.

The alpha level, df, and determination of whether the hypothesis requires a one-tailed or two-tailed test of significance are used to locate the critical value from the test. The full tables of the critical values for *r* are located in Appendix G. Below is an excerpt of the section of the r-tables that fits the current hypothesis and data. Under the conditions of an alpha level of .05, df = 8, and using a one-tailed test, the critical value is.549.

Critical Values Table

Degrees of Freedom	**one-tailed test**		
	alpha level:	**α = 0.05**	**α = 0.01**
	8	.549	.716

The critical value represents the value which must be exceeded in order to declare a result significant. The obtained value (which is called r in correlation) is the magnitude of evidence present. Because the correlation is directional and states the correlation will be positive, the obtained value must both be in the hypothesized direction (i.e. indicate a positive relationship) and exceed the critical value in magnitude. Thus, in order for the result to significantly support the hypothesis is needs to be positive and exceed the critical value of .549.

4. **Calculate the test statistic.**

A test statistic can also be referred to as an obtained value. The formula needed to find the test statistic r for this scenario is as follows:

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2}}$$

Section A: Preparation

Start each inferential formula by identifying and solving for the pieces that must go into the formula. For bivariate correlation, this preparatory work is as follows:

1. Find $\bar{X}$ (the mean for the X-variable scores)

This value is found using Data Set 12.1 and is summarized as $\bar{X}$ = 6.50

2. Find $\bar{Y}$ (the mean for the Y-variable scores)

This value is found using Data Set 12.1 and is summarized as $\bar{Y}$ = 80.00

Now that the pieces needed for the formula have been found, we can move to Section B.

Section B: Solving

The values from the preparatory steps can now be plugged into the correlation formula and used to find the r-value. Much of the work involves finding deviations, which were reviewed in detail in Chapter 4. Therefore, a summary table will be used to show deviations for the two variables in this section.

1. Find $\sqrt{\Sigma(X-\bar{X})^2}$
   a. Find $((X-\bar{X})$) by subtracting the mean of X from each x-value
   b. Find $\Sigma(X-\bar{X})^2$ by squaring each deviation and then summing those values, yielding SS_X
   c. Square root SS_X.
2. Find $\sqrt{\Sigma(Y-\bar{Y})^2}$
   a. Find $(Y-\bar{Y}$) by subtracting the mean of Y from each y-value
   b. Find $\Sigma(Y-\bar{Y})^2$ by squaring each deviation and then summing those values, yielding SS_Y
   c. Square root SS_Y.
3. Find $\Sigma(X-\bar{X})(Y-\bar{Y})$
   a. Find $(X-\bar{X})(Y-\bar{Y})$ by multiplying each deviation from X by its deviation from Y to get the product of the deviations for each case
   b. Find $\Sigma(X-\bar{X})(Y-\bar{Y})$ by summing the product of deviations for each case (i.e. sum the results from step 3a). This is the numerator for the formula.

4. Find the denominator by multiplying the square root of SS_X (which is the result of step 1c) by the square root of SS_Y (which is the result of step 2c)
5. Divide the numerator (which is the result of step 3b) by the denominator (which is the result of step 4).

Correlation Computations for Data Set 12.1.

Steps 1 Through 3							
Deviation Steps		**1a**	**1b**		**2a**	**2b**	**3a**
	Sleep Hours (X)	Deviation $(X-\bar{X})$	Dev. Squared $(X-\bar{X})$	Quiz Scores (Y)	Deviation $(Y-\bar{Y})$	Dev. Squared $(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$
	7	0.50	0.25	92	12	144	6.00
	8	1.50	2.25	88	8	64	12.00
	9	2.50	6.25	96	16	256	40.00
	6	-0.50	0.25	70	-10	100	5.00
	6	-0.50	0.25	79	-1	1	0.50
	4	-2.50	6.25	64	-16	256	40.00
	5	-1.50	2.25	75	-5	25	7.50
	10	3.50	12.25	98	18	324	63.00
	3	-3.50	12.25	53	-27	729	94.50
	7	0.50	0.25	85	5	25	2.50
Summation Steps	$\bar{X}$ = 6.50		**Step 1c** $\Sigma(X-\bar{X})$ = 42.50	$\bar{Y}$ = 80.00		**Step 2c** $\Sigma(Y-\bar{Y})$ = 1,924.00	**Step 3b** $\Sigma(X-\bar{X})(Y-\bar{Y})$ = 271.00
Step 4							
Find the denominator	$\sqrt{\Sigma(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2}=\sqrt{42.50}\sqrt{1924}=285.9545\ldots$						
Step 5							
Put the pieces together to find r and then round to the hundredths place	$r=\frac{\sum(\mathrm{X}-\overline{\mathrm{X}})(\mathrm{Y}-\overline{\mathrm{Y}})}{\sqrt{\sum(\mathrm{X}-\overline{\mathrm{X}})^2}\sqrt{\sum(\mathrm{Y}-\overline{\mathrm{Y}})^2}}$ $r=\frac{271.00}{285.9545\ldots}$ $r=.9477\ldots$ $r\approx.95$						

The obtained value for this test is .95 when rounded to the hundredths place. This is a very strong, positive correlation.

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value for r exceeds the critical value in the proper direction as follows:

a. Check the direction. The hypothesis stated the relationship would be positive. The result was positive. Thus, the direction hypothesized is supported.
b. Check the magnitude. The critical value is .549. The obtained r-value is .95. The obtained r-value exceeds (i.e. is greater in magnitude than) the critical value. Thus, the magnitude is sufficient to support the hypothesis.
c. Decide whether the evidence is sufficient to support the hypothesis

The criteria has been met for both direction and magnitude. Thus, the result significantly supports the hypothesis.

> **Note**
>
> If the hypothesis had not been directional, the only comparison needed before concluding that the hypothesis was supported would be the magnitude of the obtained r-value compared to the critical value.

6. **Calculate the effect size and/or other relevant secondary analyses.**

When it is determined that the result is significant, effect sizes should typically be computed. However, in correlation a secondary analysis is typically given rather than an effect size. The secondary computation for correlation is known as the coefficient of determination and, thus, this will be the focus of step 6 for correlation.

The **coefficient of determination** is the percent of variation in the Y-variable that is accounted for by variance in the X-variable. It can also be described as a calculation of how well a model using one variable (X) can be used to estimate the other (Y). These refer to two ways of interpreting and describing the same thing with the former more applicable to correlation and the latter more applicable to regression. We will focus on Regression in the next chapter (Chapter 13). As the focus of this chapter is correlation, we will use the interpretation language that is most applicable to correlation.

The symbol and the formula for the coefficient of determination are the same and are written as follows:

$$r^2$$

To calculate this, the obtained r-value is squared and often reported as a percent. The greatest r^2 can be is 1.00 (or 100.00% when presented as a percent). This would occur if there was a perfect correlation and could be interpreted and reported as follows:

Approximately 100.00% of the variance in Y is accounted for by variance in X.

The lowest r^2 can be is 0.00 (or 0.00% when presented as a percent). This would occur if there was no correlation (r = 0.00) and could be interpreted and reported as follows:

Approximately 0.00% of the variance in Y is accounted for by variance in X.

However, these two extreme correlation coefficients are rare and reporting a coefficient of determination is only warranted when there is a significant result for the r-value. Our result with Data Set 12.1 was r = .9477… which was significant. Thus, the coefficient of determination is warranted and would be computed as follows:

$$r^2 = (0.9477\ldots)^2$$
$$r^2 = 0.8981\ldots$$

This result is quite large and can be reported as a percent and interpreted as follows for Data Set 12.1:

Approximately 89.81% of the variance in quiz scores was accounted for by variance in hours of sleep.

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

Following this, the results for our hypothesis with Data Set 12.1 can be written as shown in the summary example below.

> **APA Formatted Summary Example**
>
> A bivariate correlation was used to test the hypothesis that hours of sleep would positively relate to quiz scores. Consistent with the hypothesis, sleep was positively related to quiz scores, $r(8) = .95, p < .05$. Approximately 89.81% of the variance in quiz scores was accounted for by variance in hours of sleep.

As always, the APA-formatted summary provides a lot of detail in a particular order. For a brief review of the structure for the APA-formatted summary of the test results, see the summary below.

Anatomy of the Evidence String

The following breaks down what each part represents in the evidence string for the correlation results in the APA-formatted paragraph above:

Symbol for the test	Degrees of Freedom	Obtained Value	r-Value
r	(8)	= .95,	r < .05.

Reading Review 12.3

1. How is df calculated for bivariate correlation?
2. What is $\bar{Y}$ and how is it calculated?
3. What information is needed to find the critical value for a bivariate correlation?
4. What does the coefficient of determination estimate?

12.7: Using SPSS

As reviewed in Chapter 2, software such as SPSS can be used to expedite analyses once data have been properly entered into the program. SPSS version 29 was used for this book; if you are using a different version, you may see some variation from what is shown here.

Entering Data

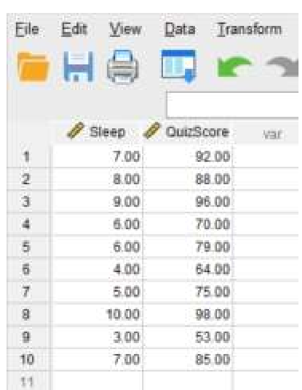
When using bivariate correlation, each variable needs to be quantitative and measured on the interval or ratio scale. If these things are all true of your data, you are ready to open SPSS and begin entering your data.

Open the SPSS software, click "New Dataset," then click "Open" (or "OK" depending on which is shown in the version of the software you are using). This will create a new blank spreadsheet into which you can enter data. Click on the Variable View tab on the bottom of the spreadsheet. This tab of the spreadsheet has several columns to organize information about the variables. The first column is titled "Name." Start here and follow these steps:

1. Click the first cell of the "Name" column and enter the name of your X-variable using no spaces, special characters, or symbols. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
2. Click the first cell of the column titled "Type" and then click the three dots that appear in the right side of the cell. Specify that the data for that variable appear as numbers by selecting "Numeric." For numeric data SPSS will automatically allow you to enter values that are up to 8 digits in length with decimals shown to the hundredths place as noted in the "Width" and "Decimal" columns, respectively. You can edit these as needed to fit your data, though these settings will be appropriate for most variables in the behavioral sciences.
3. Click the first cell of the column titled "Label." This is where you can specify what you want the variable to be called in output, including in tables and graphs. You can use spaces or phrases here, as desired.
4. Click on the first cell of the column titled "Measure." A pulldown menu with three options will allow you to specify the scale of measurement for the variable. Select the "Scale." option because the variables for a standard bivariate correlation are on the interval or ratio scale. Now SPSS is set-up for data for the X-variable.
5. Repeat steps 1-4 for the Y-variable.

Now you are ready to enter your data. Click on the Data View tab toward the bottom of the spreadsheet. This tab of the spreadsheet has several columns into which you can enter the data for each variable. The top of each column will show the names given to the variables. Click the cell corresponding to the first row of the X-variable. Start here and follow these steps:

1. Enter the data for the X-variable moving down the rows under the first column. If your data are already on your computer in a spreadsheet format such as excel, you can copy paste the data in for the variable.
2. Repeat the prior step for the Y-variable column and data. Take special care to ensure the data are case-matched. This means data for the X-variable and Y-variable for each case must share the same row.
3. Then hit save to ensure your data set will be available for you in the future. Here is how Data Set 12.1 looks after being entered into SPSS:



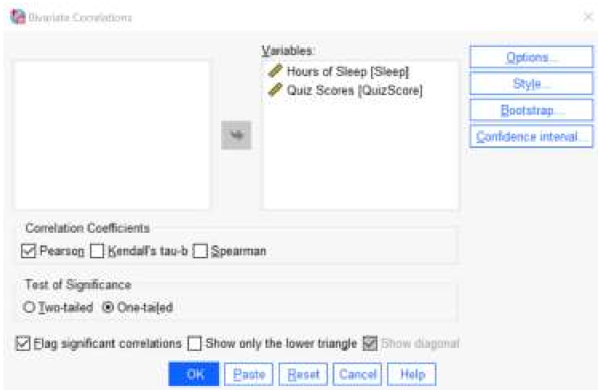
	Sleep	QuizScore	var
1	7.00	92.00	
2	8.00	88.00	
3	9.00	96.00	
4	6.00	70.00	
5	6.00	79.00	
6	4.00	64.00	
7	5.00	75.00	
8	10.00	98.00	
9	3.00	53.00	
10	7.00	85.00	
11			

Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

Conducting a Pearson’s Product Moment Correlation (PPMC) in SPSS

The steps to running this bivariate correlation in SPSS are:

1. Click Analyze -> Correlation -> Bivariate from the pull-down menus.
2. Drag the names of the two quantitative variables you wish to test from the list on the left into the Variables box on the right of the command window. You can also do this by clicking on the variable names to highlight them and the clicking the arrow to move each of them to the desired location. Select the “Pearson” option for the type of correlation coefficient. Click to select either “Two-tailed” or “One-tailed” for the test of significance for non directional or directional hypotheses, respectively. For Data Set 12.1, we would select “One tailed” because a directional hypothesis was proposed.

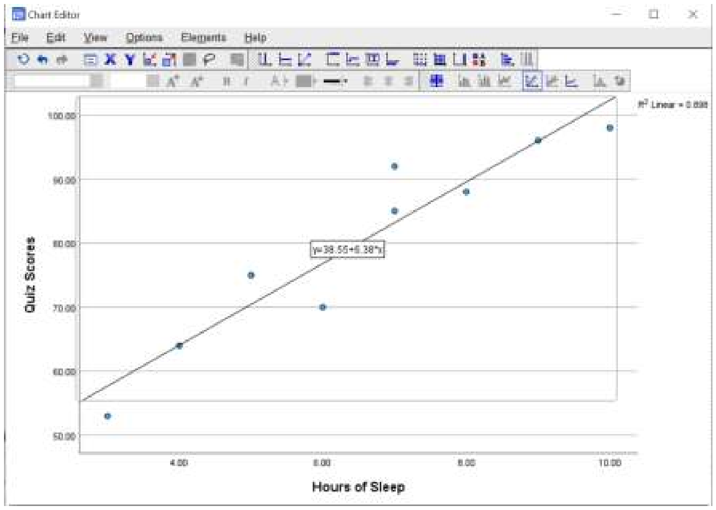


3. Click “OK” to run the analyses.
4. The output (which means the page of calculated results) will appear in a new window of SPSS known as an output viewer. The results will appear in one table as shown below.

Correlations

		Hours of Sleep	Quiz Scores
Hours of Sleep	Pearson Correlation	1	.948**
	Sig. (1-tailed)		<.001
	N	10	10
Quiz Scores	Pearson Correlation	.948**	1
	Sig. (1-tailed)	<.001	
	N	10	10
**. Correlation is significant at the 0.01 level (1-tailed).			

5. If a scatterplot is desired to visualize the correlation, click Graphs -> Scatter/Dot -> Simple Scatter -> Define. Move the names of X- and Y-variables from the left to their respective boxes on the right. Then click “OK.” The graph will appear in the output window.
6. If a fit line is desired, double-click the graph to open the graph editor. Then click the Fit Line button as shown below. Choose the “Linear” option for the fit line and click “Apply”



Reading SPSS Output for a Bivariate Correlation

The table shows the correlation coefficient on the first line which is r = .948. This is .95 when rounded to the hundredths place for reporting purposes. This is a very strong, positive correlation. The second line shows the chance of a Type I Error which is the p-value; SPSS calls this value "Sig." The output show that for Data Set 12.1, the p-value is < .001. The alpha level chosen was .05 so this result is significant because the p-value is less than the alpha-level of .05 (i.e. $p < .05$). The third line shows the sample size but SPSS uses the symbol N rather than n. The output show that the sample size for Data Set 12.was 10. These match the results from the hand-calculations performed earlier in this chapter for Data Set 12.1.

The scatterplot with the fit line shows a positive slope. The coefficient of determination is shown in the top right corner of the graph as r^2 = .898. This means that approximately 89.8% of the variance in Quiz Scores was accounted for by Hours of Sleep. This also matches the result from the hand-calculations performed earlier in this chapter for Data Set 12.1.

Reading Review 12.4

1. What scale of measurement should be indicated in SPSS for both variables when performing a Pearson's correlation?
2. What information is provided in each of the three lines of SPSS output for correlation?
3. Which kind of graph should be used for a bivariate correlation?
4. How can the coefficient of determination be checked using SPSS?

12.8: Structured Summary for Bivariate Correlation

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes.

Summarize what each symbol stands for.

n =

df =

$\bar{X}$ =

$\Sigma(X-\bar{X})^2$ =

$\bar{Y}$ =

$\Sigma(Y-\bar{Y})^2$ =

$\Sigma(X-\bar{X})(Y-\bar{Y})$ =

Fill-in the appropriate information for each section below:

1. Bivariate Correlation Basics
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What assumptions must the data meet to use this test?
2. Bivariate Correlation Formula
   a. What is the formula for a Bivariate Correlation?
   b. What things should be computed in the preparatory steps for using this formula?
   c. What are the steps for solving using this formula?
3. Reporting Results from a Bivariate Correlation
   a. How is this statistic reported when using APA format?
      i. What things must be reported in the APA summary sentence for bivariate correlation?
      ii. What specific additional statistic is often reported and interpreted as a percent when the correlation is significant?

Bivariate Correlation Calculations Chart

Steps 1 Through 3							
Deviation Steps		**1a**	**1b**		**2a**	**2b**	**3a**
	Sleep Hours (X)	Deviation $(X-\bar{X})$	Dev. Squared $(X-\bar{X})^2$	Quiz Scores (Y)	Deviation $(Y-\bar{Y})$	Dev. Squared $(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$

Summation Steps	$\bar{X}$ =		**Step 1c** $\Sigma(X-\bar{X})^2$ =	$\bar{Y}$ =		**Step 2c** $\Sigma(Y-\bar{Y})^2$ =	**Step 3b** $\Sigma(X-\bar{X})(Y-\bar{Y})$ =
Step 4							
Find the denominator	$\sqrt{\Sigma(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2} =$						
Step 5							
Put the pieces together to find r and round to the hundredths place	$r = \dfrac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2}}$ $r =$						
Additional Analyses							
Compute df for all correlations Computer r^2 only for significant correlations	$df = n-2$ r^2 = the coefficient of determination						

CHAPTER OVERVIEW

13: Simple Linear Regression

13.1: Introduction to Linear Regression

In its most basic form, **linear regression** is a technique built upon correlation to test whether and how well the values from one variable can be used to estimate values for another variable. Because this builds from the ideas used in correlation, it is important to review Chapter 12 before starting this chapter. When a bivariate correlation is appropriate to the data but the hypothesis is about prediction or estimation, a linear regression is the best fit. The version of this which is the focus of this chapter can be referred to by several names, including: simple regression, linear regression, bivariate linear regression, or bivariate regression using a least squares model.

The simplest form of regression is a bivariate (i.e. two variable) form called simple linear regression. Simple linear regression establishes whether there is a relationship between two quantitative variables and, if so, uses one to estimate the other. In this way, you can think of linear regression as a companion to a correlational analysis. For example, a bivariate correlation could be used to test whether income is related to level of happiness or whether hours spent exercising are related to amount of stress. A regression would be used to test whether income is (mathematically) predictive of happiness or whether hours of exercise are useful in estimating level of stress.

13.2: Variables

Just as was true for bivariate correlation, simple linear regression can be used when you have two quantitative variables. Each quantitative variable should be continuous and measured using an interval or ratio scale of measurement. The two variables are often referred to as X and Y when spoken about generally. The scores for these variables are referred to as X-values and Y-values, respectively.

13.3: Data and Assumptions

Each statistical test has some assumptions which must be met in order for the formula to function properly. In keeping, there are a few assumptions about the data which must be met before a simple linear regression is used. Note that these are the same assumptions that must be met for bivariate correlation. First, the scores for both variables must be matched. This means that the same participants have scores for each variable and that the two scores for each participant can be identified together. Second, data for each quantitative variable should be fairly normally distributed without notable impact due to either univariate or bivariate outliers. A bivariate outlier refers to a participant whose scores on X and Y together do not follow the general pattern of the other participants; if their data diverge markedly from the general pattern, they are considered a bivariate outlier. Finally, the relationship between the variables should be linear meaning they approximate a straight line rather than other shapes such as would be seen when graphing a quadratic equation, for example. See Chapter 12 for a review of how to assess a linear pattern visually using scatterplots.

13.4: Prediction in Regression

Prediction in regression is the idea that if two things are related, one can be used to estimate or approximate the other but this does not mean one causes the other or that we are mathematically seeing into the future. The idea is much simpler than that. When we use prediction in regression we are starting by establishing a pattern in data and then saying, essentially, "If a pattern is generally true based on known data points, it can be useful in estimating new or previously unknown data points." For example, suppose that in data collected from 100 people, there is a very strong, positive relationship between hours spent studying and exam scores. This would mean that, among those participants, exam scores tended to increase as hours of studying increased. This would also mean that as time spent studying decreased from person to person, exam scores also tended to decrease. Based on this pattern, if someone new reported they did not study at all we would predict that they would have a relatively low score on the exam. In keeping, if someone new reported they studied for many hours we would predict that they would have a relatively high score on the exam.

The Meaning of "Prediction"

It is important to clarify what prediction does and does not mean when this term is used with regression. When regression is used, **prediction** refers to the estimated, expected value of one variable (Y) based on the known value of another variable (X). Note that in this definition the prediction is "estimated" and "expected" rather than "precise" and "known." The terms prediction and estimation may be used interchangeably for regression.

Regression Lines

Correlations and regressions summarize relationships between two quantitative variables using a line with a consistent slope. When this is done with correlation, the summary line is often called a fit line but when regression is used the summary line is called a regression line. A **regression line** represents the best approximation of the linear relationship between two variables using a fit method, the most common of which is the least-squares regression method. A *fit method* refers to the way the line was created to best fit the data. There are different methods that can be used for creating the regression line. One is the least squares model. This fit method creates the regression line by angling it such that it minimizes residuals based on the data provided. Therefore, regressions using this are often called least-squares regression models. However, this method is so common and useful that it is often assumed when the form of regression is not specified. This is the version of regression which will be used throughout this chapter.

Least square regression refers to when the best-fitting line for the data is calculated such that the sum of squared deviations from Y (i.e. the vertical deviations) is minimized. To state this another way, this procedure is used to find the exact angle for the line that gets as close to the data on the graph (i.e. the dots on a scatterplot) as possible, on average. This should result in as many data points falling above the line as below the line. The line will only be able to actually pass through every data point when the correlation is perfect, which is quite rare. Instead, therefore, the line summarizes the pattern among the data such that the distance between the dots and the line is minimized when a correlation exists but is not perfect.

Consider what we already know about the fit lines for correlation in Chapter 12 (which are the same as the regression lines in this chapter). The stronger a relationship is between two variables, the closer the data points will tend to be to the line. Conversely, the weaker the relationship is between the two variables, the farther the dots will tend to be from the line. When the relationship is stronger, the line is a better fit to the data and when it is weaker, the line is a poorer fit to the data.

The Basics of β_1 (slope of a Regression Line)

Predictions in regression use the slope of the regression line plus something known as the y-intercept. For this reason, understanding the slope is key to understanding and using regression. The slope of the regression line summarizes the pattern between the X-variable and Y-variable. The symbol for the slope of a regression line is β_1. This symbol is the Greek letter beta. Thus, this symbol with the subscript of 1 is known as "beta one." We are only testing one regression and one slope at a time in this chapter so we will only be considering beta one. However, more complex techniques beyond the scope of this book may test multiple slopes at once. Additional slopes (i.e. betas), when applicable, are numbered consecutively starting from 2.

The slope of a line (β_1) is computed as change in the Y-variable divided by corresponding change in the X-variable. The symbol for change is Δ. Slope is often summarizes as "rise over run" meaning vertical change (i.e. y-axis change) divided by horizontal change (i.e. x-axis change). In mathematics the slope is, thus, often written as a division problem. Statisticians use the quotient of that division problem for the slope known as β_1. Thus, slope can be presented as a fraction or as its quotient but is usually reported in quotient form in statistics.

$$\beta_1 = \frac{\Delta Y}{\Delta X}$$

When β_1 is used (in quotient form), it indicates the change in the Y-variable predicted for each one unit increase in the X-variable. Thus, if β_1 = 0.30 it translates to saying that for every one unit increase in the X-variable, there is a .30 unit increase predicted in the Y-variable. However, if β_1 = -0.30 it translates to saying that for every one unit increase in the X-variable, there is a 0.30 unit decrease predicted in the Y-variable.

When a hypothesis is tested, the generic terms of X-variable and Y-variable are replaced with the actual variable names of variables being tested when interpreting a slope. Suppose that a hypothesis is tested which states that "Hours spent studying will predict exam scores" and that when the data are tested, it is found that β_1 = 7.00. This slope would translate to saying that, "For every one hour increase in studying, there is a 7.00 unit increase predicted for exam scores." Another way to word this is to say that, "For every additional hour spent studying, a 7.00 point increase in exam scores is predicted."

13.5: Hypotheses

Hypotheses for regression can be directional or they can be non-directional. Directional hypotheses require a one-tailed test of the hypothesis and non-directional hypotheses require a two-tailed test of the hypothesis. Regression is used to assess whether one variable is useful in predicting another and, if so, in which way and how accurately. As noted, the predictions and the direction of the relationship are estimated and summarized using the slope (β_1). When β_1 is positive, the slope is positive. When β_1 is negative, the slope is negative. When β_1 is 0, it means that there is no discernable slope and that the variables are unrelated. Thus, directions are interpreted with slopes in regression in a similar way to how directions are interpreted with r-values in correlation.

We will focus on the non-directional hypothesis for this chapter. For simple linear regression, the non-directional research hypothesis is that the X-variable will be a significant predictor of the Y-variable. This means that the slope will not be 0. The corresponding null hypothesis is that the X-variable will not be a significant predictor of the Y-variable. The null hypothesis, therefore, states that the slope will be 0. The non-directional research and corresponding null hypotheses for simple linear regression can be summarized as follows:

Non-Directional Hypothesis for Simple Linear Regression

Research hypothesis	Variable X will predict Variable Y.	$H_A : \beta_1 \neq 0$
Null hypothesis	Variable X will not predict Variable Y.	$H_0 : \beta_1 = 0$

Keep in mind that we estimate and approximate in statistics. Therefore, the slope does not have to be exactly 0 to retain the null. Instead, it is presumed that the slope should be treated as 0 and that the null hypothesis should be retained unless the results are statistically significant. We will see how this is visualized and significance is tested in the remaining sections of this chapter.

Reading Review 13.1

1. What does prediction mean in regression?
2. Which variable is being predicted in regression?
3. How is a regression line created?
4. How would β_1 = 0.75 be interpreted?
5. How would β_1 = -2.50 be interpreted?

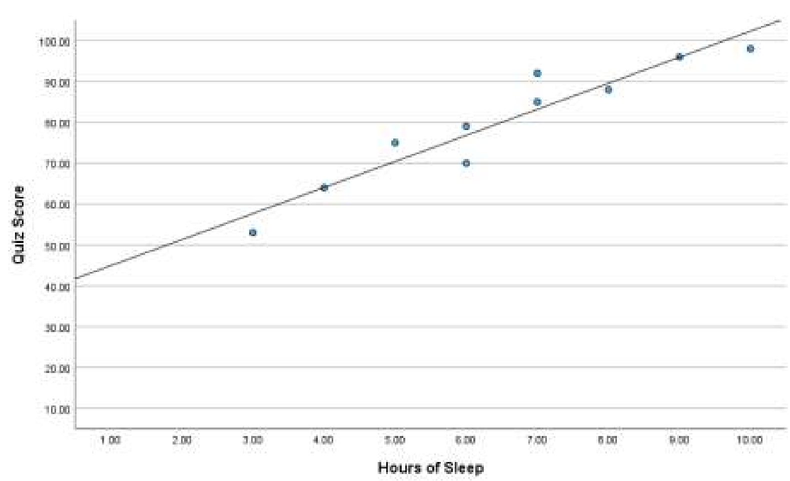
13.6: Visualizing Linear Regression

Regression builds for the same graph and foundation as correlation. Thus, bivariate regression can be graphed using a scatterplot and a regression line (See Chapter 12 for a review of scatterplots and fit lines). The **regression line** is used to summarize the change in Y that is associated with a change in X. The closer the dots tend to be to the regression line, the stronger the relationship and, thus, the more accurate the predictions will be based on the regression line. When there is a relationship between the variables, a regression line will either slope up or it will slope down when read from left to right. The angle of the line (i.e. the slope) indicates the direction of the correlation (either positive or negative).

Let’s take a look at how regression works using Data Set 12.1 from Chapter 12. In Chapter 12, we supposed a researcher collected data from 10 college students to test the hypothesis that hours of sleep would positively relate to quiz scores. For regression we would hypothesize that hours of sleep would be useful in predicting quiz scores. Data Set 12.1 and the corresponding scatterplot for those data are included below for reference.

Data Set 12.1. Hours of Sleep and Quiz Scores (n = 10)

Participant Number	Sleep Hours	Quiz Score
1	7	92
2	8	88
3	9	96
4	6	70
5	6	79
6	4	64
7	5	75
8	10	98
9	3	53
10	7	85



Graph 12.2 Hours of Sleep and Quiz Scores with a Regression Line

Predictions using Regression Lines

The stronger the correlation, the more accurate the predictions. This is because the regression line is being used to make predictions. Recall that a regression line is balanced to approximate the location of the dots with as little error as possible. Keep in mind that each dot is a bivariate data point. When we look at those data points in Graph 12.1, we can see that they are fairly close to the positively sloping regression line. Visually, we can see that the line does a good job of estimating the location of the data and, thus, that it will likely be useful in estimating (predicting) Y-values using X-values.

The Regression Equation

When regression is used, the regression line is being used to estimate scores of the Y-variable. Thus, the equation of the line is the formula used to predict (or estimate) Y-values. Linear graphs are summarized with the following equation:

$$\hat{Y} = b_0 + b_1 x$$

In this version of the linear equation $\hat{Y}$ stands for a predicted Y value, b_0 stands for the y-intercept of the line, and b_1 stands for the slope of the line. A y-intercept is where the line crosses the y-axis; it represents what Y equals when X is zero. Notice that this is very similar to the structure some of us have seen before in math classes when we learned the linear equation; the version used in math classes is as follows: $Y = mx + b$. The symbols for the slope and y-intercept are different in statistics and math but they represent the same things. Specifically, in math the slope is called "m" and the y-intercept is called "a" while in statistics the slope is called "b_1" and the y-intercept is called "b_0". In statistics, the y-intercept, b_0, is also often referred to as a constant because it refers to a single point on the graph. In addition, in statistics we often state the y-intercept first followed by slope times X whereas in math we may see these reversed (where the slope times X is stated first and the y-intercept is then added). However, these are just two different ways of writing the same formula. Thus, the two formulas are computationally the same, even though the symbols look different.

The regression equation is what is used to both summarize the regression line and to make predictions. When the slope and y-intercept are known, an X-value can be plugged into the linear equation to predict a Y-value. Slope and y-intercept are rarely calculated by hand in regression and, instead, are generally calculated using software such as SPSS. Therefore, the process for hand-calculating the slope and y-intercept are not included in this chapter and, instead, we will focus on how to read and use these values using SPSS as we progress through this chapter.

13.7: The Purpose of the Four Parts of a Regression Analysis

So far we have reviewed the types of hypotheses and data which are a good fit to regression, what prediction means in regression, and the logic behind using the regression line to predict a Y-value. However, a regression line should only be used to make predictions when it is sufficiently useful in doing so. Thus, we need to understand how to test a hypothesis to know *if* X is useful for predicting Y before we use it to do so.

Regression is a complex technique which uses three sets of analyses to test a hypothesis and establish whether X is useful in predicting Y. These three sets of analyses are: Correlation, ANOVA, and a *t*-Test. When it is determined that X significantly predicts Y, a fourth component can then be used to make predictions. This component is the linear equation of

$$\hat{Y} = b_0 + b_1 x$$

Thus, regression is actually a technique that draws from other existing techniques and puts them together to serve the new purpose of predicting.

It can be helpful to have a broad idea of the major parts of regression and their purposes before going into details about each. Therefore, we will start with a brief overview of the role of each of the components of regression before going into detail about how each is computed.

The four components are:

1. Correlations to establish whether X relates to Y,
2. ANOVA to test whether using X significantly reduces error in predicting Y,
3. *t*-testing to assess whether it is the slope of the line that is reducing the error in predictions and, when warranted,
4. The linear equation to make predictions.

Each of these four components answers a different question we are posing when using regression.

Correlation in Regression

Correlation is used in regression to answer the question:

Does X relate to Y?

Statisticians will often check whether there is a significant relationship between an X-variable and a Y-variable before progressing to testing predictions. When X relates to Y, the coefficient of determination may also be computed and reported as part of a test using regression.

Importantly, the scatterplot with the fit line from correlation is a visual depiction of a regression model (Recall that a fit line in regression is called a *regression line.*). "Regression model" simply refers to how X is being used to predict Y with the linear equation. The regression line is the foundation from which predictions are made in regression. Thus, when the correlation is significant between X and Y, it offers support for a regression model which uses X to predict Y. For a detailed review of correlation, see Chapter 12.

Though correlations are often checked in regression, they are not always reported for bivariate regression. This is because the results of the correlation are redundant to the results of the ANOVA and *t*-test when there is only one X-variable (predictor) being tested in a regression model. Instead, the coefficient of determination (r^2) is often checked and reported for simple regression and r can be checked but does not generally need to be reported.

ANOVA in Regression

ANOVA is used in regression to answer the question:

Does using X significantly improve predictions of Y?

ANOVA is a very important part of a regression because it is used to test the central aspect of the hypothesis: whether X predicts Y. The ANOVA portion of regression is used to compute how good X is at predicting Y by assessing whether it significantly reduces the error of predictions compared to an alternative prediction model. Thus, improvement in predictions is defined and estimated as reduction in errors in predictions when using the regression model. By *regression model* we simply mean a model where X is used to predict Y. The greater the reduction of error when using X to predict Y, the more useful the regression model is and, thus, the larger the ANOVA F-value will be.

The regression model must be compared to an alternative model. The default alternative way to predict any Y-value is to simply use the mean of Y for all predictions. The symbol for the mean of Y is $\bar{Y}$. When the regression ANOVA is significant, it indicates that using X significantly improves predictions of Y over using $\bar{Y}$ as the prediction for all Y-values. Thus, when the ANOVA is significant, it offers support for a regression model which uses X to predict Y. This concept is central to testing a regression hypothesis and, thus, we will focus on this in quite a bit of detail in this chapter.

t-Test in Regression

A *t*-test is used in regression to answer the question:

Does the slope of the regression line significantly improve the predictions of Y?

The slope of the regression line is what allows the line to get as close to all the data points as possible, on average. Regression models use one or more X-variables to predict a Y-variable. For this chapter we are focused on bivariate regression where there is only one X-variable. However, when multiple X-variables are used to predict Y (which is common in research), they will each contribute a different slope. In regression, *t*-tests are used to assess which of those slopes significantly improved predictions and which did not. Think of this as the post-hoc part of a regression; if the regression ANOVA is significant, the *t*-tests are used to see which X-variables had slopes that significantly contributed to the accuracy of predictions. The improvements gained by the slope of each X-variable are tested with separate *t*-tests. However, in bivariate linear regression, only one X-variable is being tested meaning only one slope is being used to predict Y. In this case, if the ANOVA is significant, the *t*-test will also be significant. Therefore, assessing the significance of the *t*-test is less essential for a bivariate regression than in multivariate regression. However, it should still be assessed and reported as part of a complete regression analysis.

In addition, the slope and y-intercept for the regression formula are computed when using SPSS and can then be reported with the *t*-test results. Thus, we will review how to read and interpret results for the *t*-test component of regression to get necessary information about the slope and the y-intercept.

Making Predictions in Regression

When a regression is significant, it establishes that using the regression line and X-values are useful in predicting Y. It follows then that the regression equation should be used to predict Y when regression results are significant. Thus, the last part of the regression is interpreting how the slope predicts Y and creating the corresponding regression equation which can be used to predict Y. Predictions of Y are made using the linear equation as follows:

$$\hat{Y} = b_0 + b_1 x$$

In this equation, a predicted Y value ($\hat{Y}$) is computed by multiplying the slope of the regression line (b_1) by an X-value (x) and then adding the y-intercept (b_0). Slopes and intercepts can be computed by hand or using SPSS. Later in this chapter, we will review how to find and interpret these using SPSS. In addition, slopes are typically interpreted and included as part of a complete APA-formatted results paragraph for regression. The regression equation can also be constructed and used to predict Y-values, if desired, but it is not always included in a results paragraph.

1. What is correlation used to check in a regression?
2. What does ANOVA compare when used to test a regression model?
3. What part of the regression model is tested by the *t*-test in a regression?
4. Under what conditions and for what purpose is the linear equation used with regression?

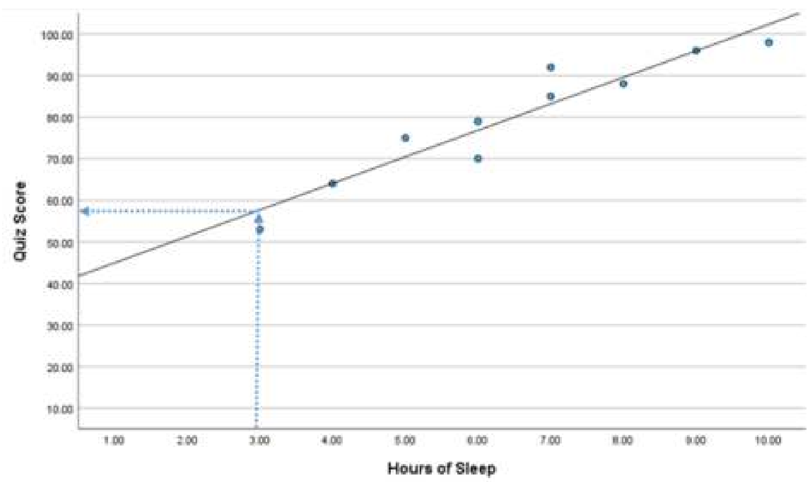
13.8: Visualizing Predictions and Residuals in Regression

ANOVA is used to test whether the regression model is a good fit to the data and, thus, is useful for predicting Y. It does this by computing and comparing the errors in predictions when using the regression model to an alternative model using the mean of Y. The computations can be easier to understand if we can visualize what each model does and what their errors represent before learning how to compute the corresponding parts of the ANOVA. Thus, this section will review how each model can be represented in graphical form.

Visualizing Predictions Using the Regression Equation

Predictions are made based on the regression line. The regression line summarizes where scores are expected to be. Later in this chapter we will learn how to precisely calculate the predicted Y-value. However, for this section we are reviewing how to estimate values visually to help us understand the logic of regression. With that in mind, let's take a look at an example using Graph 13.1 below. Suppose we wanted to predict the exam score for someone who slept 3 hours.

To locate the prediction on the graph, we would look over to 3.00 on the x-axis because that corresponds to getting 3 hours of sleep (i.e. 3 units for the predictor variable). We would then find where the regression line crosses over X = 3.00; the height of the regression line where it crosses over X = 3.00 indicates the predicted Y-value. We can see in the graph that when X = 3.00, Y is around 58 units high on the y-axis. This is the approximate predicted score for someone who gets 3.00 hours of sleep. To state it another way, someone who gets 3.00 hours of sleep is predicted to get a quiz score of *approximately* 58. Notice the imprecision of the estimate when looking visually. This is why a formula must be used later to get a more precision. For now, however, we will stick to the visual estimate.

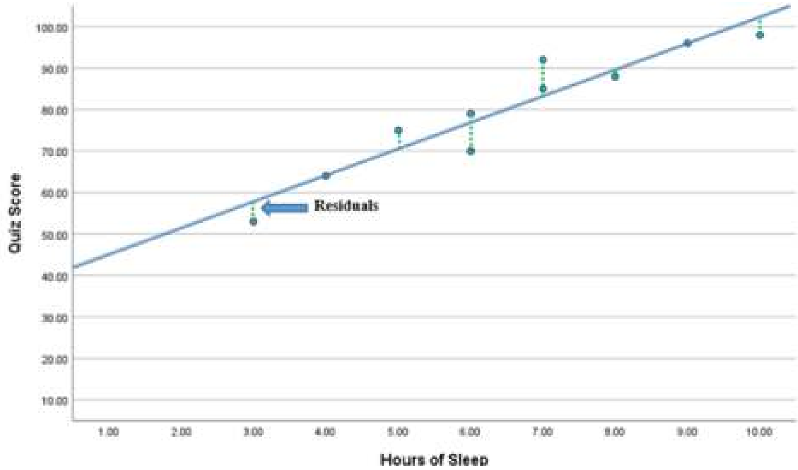


Graph 13.1 Regression Line Used to Predict Y when X = 3.00.

We can compare our predicted Y-value to an actual, known Y-value. Specifically, we can see that there is a data point for which X was 3.00. That data poi falls at X = 3.00 and Y = 53.00. This data point is a known value from Data Set 12.1 and corresponds to participant number 9. Notice that the known data point is not on the regression line. It is close to the prediction (i.e. it is close to the regression line) but not exact which means that the prediction was pretty good but did have some error. This error is known as a *residual* and is an important part of estimating how accurate and, thus, how useful a regression is for making predictions.

Residuals Using the Regression Line

Residuals represent the amount of *inaccuracy* in the regression predictions. Specifically, **residuals** are the errors in locating actual Y-values when using the regression line and represent the vertical distances between the known bivariate data points and the regression line. Another way say this is that residuals represent the amount of variation in the Y-variable that is not accounted for using the X-variable. Below is Graph 13.2 with the residuals shown for all data points. Notice that the residuals are all shown vertically. This is because we are assessing error in predicting Y and, thus, are concerned with error in locating the data on the y-axis (which is vertical). Recall from Chapter 12 that the correlation between sleep hours and quiz scores for Data Set 12.1 was summarized as r = .95 (See Chapter 12 for review of the computations). This is a very strong, positive correlation. When a correlation is stronger, the average residuals will be lower. Consistent with this, the residuals for Graph 13.2 are fairly small, overall, as we should expect because the correlation was very strong. Thus, predictions of Y using X will be fairly accurate (though imperfect) for Data Set 12.1.

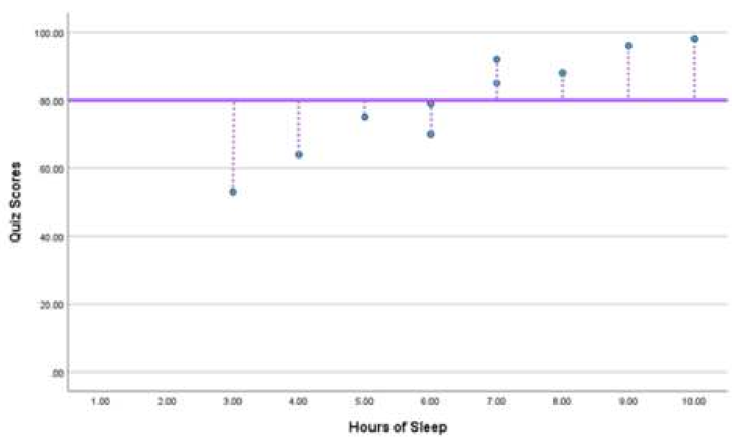


Graph 13.2 Regression Line with Residuals Shown.

Visualizing the Alternative Prediction Model: Using $\bar{Y}$ as the Prediction

To know whether using X to predict Y significantly improves predictions, it needs to be compared to another model. The alternative way to predict Y is simply to predict that all Y-values are equal to the mean of Y. This alternative is based on the fact that means are summaries of what tends to be true of data for a given variable. When data are normally distributed, scores closer to the mean are more common and scores farther from the mean are rarer. Thus, it is logical to use the summary of what tends to be true (which is the mean for the variable) as an alternative way to estimate values.

Let's take a look at the graph using this alternative prediction model (see Graph 13.3). The mean of Y is 80.00 (i.e. $\bar{Y}$ = 80.00) for Data Set 12.1. The prediction line is, therefore, set as $\hat{Y}$ = 80.00 and all Y-values are predicted to be 80.00. This is represented by a non-sloping, horizontal line. Just as we did with the regression line in Graph 13.2, we can draw vertical residual lines from each data point to the prediction line (which in this case is the $\bar{Y}$ line) to represent error in predictions. The alternative model residuals are depicted in Graph 13.3.

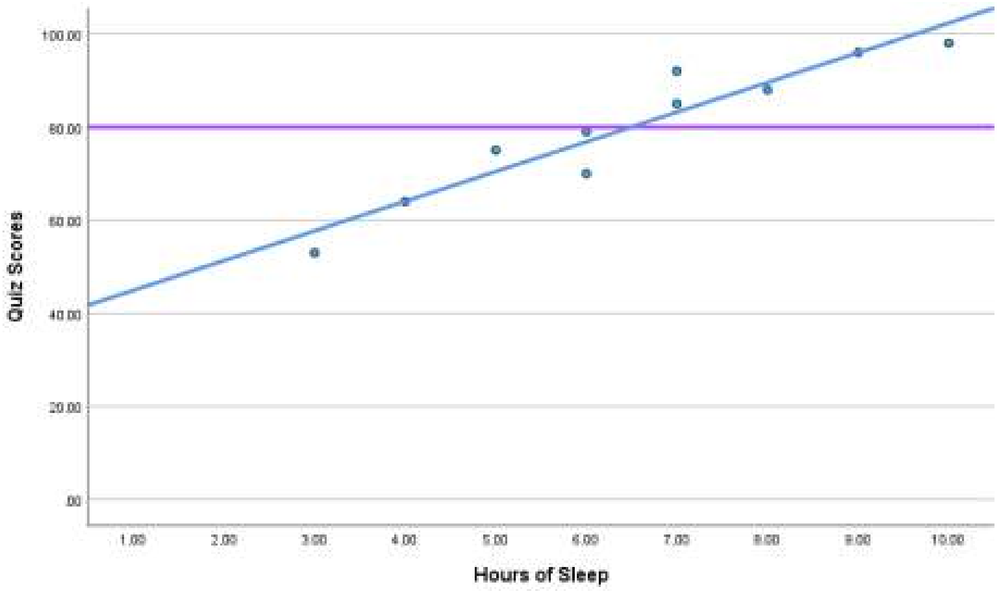


Graph 13.3 Mean of Y Line with Residuals Shown.

Visually Comparing Model Fit: Comparing Predictions with the Regression Line vs. the $\bar{Y}$ Line

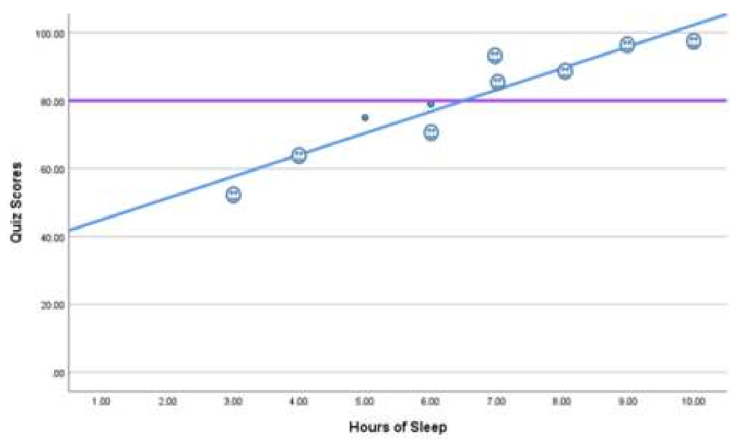
Model fit refers to how good a model is at making predictions. It is called *model fit* because we are asking how well the model fit the data. For regression this means we want to know how close the prediction lines were to the actual data. The lesser the residuals, the better the fit and the greater the residuals, the poorer the fit. The best way to assess the fit of a regression model is by pitting it against another prediction model. This is precisely what is done in simple linear regression. The residuals from two models are compared: 1.The model predicting Y using X (the regression model) and 2. The model predicting each Y is equal to the $\bar{Y}$.

Let's compare the two models visually to see which is a better fit to the data and, thus, is more useful for predicting Y-values. Graph 13.4 shows both models and their residuals. This version overlays the regression line (which is blue and has a positive slope) to the $\bar{Y}$ line (which is purple and has no slope). Whichever model's line is closer to a data point is the one which is better at predicting that data point.



Graph 13.4 Comparison of Prediction Models

Let's make it more overt by marking how many of the data points were closer to the regression line than the $\bar{Y}$ line with a smiley face. Because our hypothesis is that the regression line will be the better predictor, Graph 13.5 shows smiling faces over the dots which are better predicted by the regression line than the alternative model. We can see that for 8 of the 10 data points, the regression model was better. This means the residuals were lower using the regression model for 8 of the 10 data points than when using the alternative model. For one of the remaining data points, the two models were about equally accurate in their prediction and for the other the alternative model was more accurate. Overall, we can see that the residuals are lower for the regression model, in general, than for the alternative model. Thus, though the regression model was not always more accurate, it tended to be the better model for predicting.



Graph 13.5 Data Points Better Predicted by the Regression Model than $\bar{Y}$

In this section we have taken the time to visually understand:

1. the two models used to predict Y-values,
2. what residuals represent in the models, and
3. how the two models are being compared for fit based on their residuals.

In a regression, the residuals of the two models are calculated and used to test whether the regression model provides significantly improved predictions over the alternative model. We will now to turn to those calculations. You may find it helpful to refer back to the visuals in this section as we cover those calculations to remind yourself what is being represented in each part of the regression ANOVA.

1. How is a predicted Y-value found on the graph of the regression model?
2. What does a residual represent when using a regression line to predict a Y-value?
3. How is a predicted Y-value found in the alternative prediction model?
4. Which model has a sloping prediction line?
5. What is compared to test whether the regression model improves predictions compared to the alternative model?

13.9: Testing a Regression Model

Testing a Regression Model with ANOVA

ANOVA is used to test whether the regression model is a good fit to the data and, thus, is useful for predicting Y. To do so, three kinds of error are computed:

1. Residuals using the alternative model (i.e. the $\bar{Y}$ model),
2. Residual using the regression model (i.e. using $\hat{Y} = b_0 + b_1 x$), and
3. Residuals in the alternative model that are not accounted for using the regression model.

Residuals using the alternative model are known as the sum of squares total (SST). Think of SST as the error we would have if we used the simplest way of predicting. If the regression model is useful, it will produce residuals that are significantly lower than the SST. Thus, we must also compute the residuals using the regression model. The residuals that occur when using the regression model are known as the sum of squares errors (SSE); SSE summarizes the residuals that are not accounted for using the regression model. The difference between the residuals for the alternative model (SST) and the regression model (SSE) are the residuals that have been accounted for using the regression model; these are known as the regression sum of squares (a.k.a. sum of squares regression; SSR).

The three forms of residuals and what they stand for can be summarized as follows:

Error in the Alternative Model = Error Explained Using the Regression Model + Error Left Unexplained Using the Regression Model

SST = SSR + SSE

Think of SSR as the amount of error that is reduced when using the regression model. The higher the SSR, the lower the SSE and the better the regression model is at predicting Y-values.

Computing Residuals for the Alternative Prediction Model: Sum of Squares Total (SST)

Residuals represent errors in predictions. The overall error observed is computed as the residuals when using the alternative model (SST). In this model, the mean of the Y-values is used as the prediction for all data points, regardless of their X-values.

Recall that residuals are differences between the predicted Y-values for a data set and the actual Y-values for the data set. The predicted Y-values in the alternative model are all the mean of Y. Therefore, a horizontal line at the $\bar{Y}$ is what is used to predict in the alternative model. As we can see in Graph 13.3, the $\bar{Y}$ prediction line is too high for some data points (i.e. the predicted Y is higher than the actual Y value) and is too low for other data points (i.e. the predicted Y is lower than the actual Y value). Thus, some residuals are negative while others are positive. Because the $\bar{Y}$ is being used, which balances deviations so that they sum to 0, it also causes residuals to balance so they sum to 0. To avoid this issue when computing total residuals, a squared version is used. This use of squaring to is just like we have saw when working with standard deviations in earlier chapters (such as Chapter 4).

The **sum of squares total (SST)** in regression refers to the deviations when using the alternative model. To compute SST, these four steps are followed:

1. Find $\bar{Y}$ (the mean of the Y-values)
2. Find the residuals by subtracting the $\bar{Y}$ from each known Y-value
3. Square the residuals
4. Sum the squared residuals to get the SST

Let's compute the SST using Data Set 12.1. In Table 13.1, we see the data and three additional columns:

1. The predicted quiz scores (i.e. the predicted Y-values),
2. The residuals, and
3. The squared residuals.

Notice that $\bar{Y}$ is used for all predicted scores in the alternative model. The sum of squared residuals (SST) is shown at the bottom of the table. The SST is 1,924.

Table 13.1: Computing Residuals for Data Set 12.1 Using the Alternative Model

Sleep Hours	Actual Quiz Scores	Predicted Quiz Scores ($\bar{Y}$)	Residuals	Squared Residuals
7	92	80	12	144
8	88	80	8	64
9	96	80	16	256
6	70	80	-10	100
6	79	80	-1	1

Sleep Hours	Actual Quiz Scores	Predicted Quiz Scores ($\bar{Y}$)	Residuals	Squared Residuals
4	64	80	-16	256
5	75	80	-5	25
10	98	80	18	324
3	53	80	-27	729
7	85	80	5	25
				SST = 1,924

Computing Residuals for the Regression Model: Sum of Squares Error (SSE)

The hypothesis when using regression is that the residuals will be significantly lower when using the regression model than the alternative model. Therefore, we must also compute the residuals when using the regression model. This can be thought of as the error that is left unaccounted for, unreduced, or unexplained when using the regression model. The error that occurs when using the regression model is known as the sum of squares error (SSE). It is important to note that SPSS software refers to SSE as "Sum of Squares Residual" rather that "Sum of Squares Error" (see Table 13.3 for a summary of the various names and symbols used for residuals computations in regression).

Before we can compute the residuals, we must make predictions using the regression model. The regression formula used to predict Y-values is:

$$\hat{Y} = b_0 + b_1 x$$

To use this formula to make predictions, we must first know the y-intercept (b_0) and the slope (b_1) for the regression line. Computing these by hand is beyond the scope of this chapter. Instead, these are often generated using software such as SPSS. Thus, we will use the slope and y-intercept as computed in SPSS. For Data Set 12.1, the slope and intercept are as follows:

b_0: Y-intercept	38.5529
b_1: Slope of Hours of Sleep	6.3765
Note: Values continue but are rounded to the fourth decimal place for space.	

Thus, the predicted Y-values for the regression model with Data Set 12.1 are computed using the following regression equation:

$$\hat{Y} = 38.5529 + 6.3765x$$

To use this equation, the X-value for each case is plugged in. It is then multiplied by the slope. Finally, it is added to the y-intercept. For example, in the first case for Data Set 12.1, the X-value is 7. Thus, the predicted Y-value for that case is computed as follows:

$$\begin{gathered}\hat{Y} = 38.5529 + 6.3765(7) \\ \hat{Y} = 38.5529 + 44.6355 \\ \hat{Y} = 83.1884\end{gathered}$$

This process is repeated to find the $\hat{Y}$ for each case. The $\hat{Y}$ for each case is shown in Table 13.2.

Now that the predicted Y values are known, the sum of squares errors can be computed. **Sum of squares errors (SSE)** is computed by finding the residuals for each data point when using the regression model, squaring those residuals, and then summing them to get a total. Thus, the summary of the steps to compute SSE are as follows:

1. Find the predicted Y-values for each case using the regression equation, which is: $\hat{Y} = b_0 + b_1 x$
2. Find each residual by subtracting the $\hat{Y}$ from each known Y-value
3. Square the residuals
4. Sum the squared residuals to get the SSE

Let's compute the SSE using Data Set 12.1. Table 13.2 includes the data in the first two columns and three additional columns of computations:

1. The predicted quiz scores (i.e. the predicted Y-values),
2. The residuals for those predictions, and
3. The squared version of those residuals.

Notice that the predicted Y-values vary because each depends upon the corresponding X-value for the case. The sum of squared residuals (SSE) is shown at the bottom of the table. The SSE is 195.9766. This value represents the error that occurs when using the regression model to predict. Notice that this is much lower than the error we saw when using the alternative model to predict (i.e. the SSE of 195.9766 is noticeably lower than the SST of 1,924). This is a desirable outcome.

Table 13.2: Computing Residuals for Data Set 12.1 Using the Regression Model (SSE)

Sleep Hours	Actual Quiz Scores	Predicted Quiz Scores ($\hat{Y}$)	Residuals	Squared Residuals
7	92	83.1882	8.8118	77.6473
8	88	89.5647	-1.5647	2.4483
9	96	95.9412	0.0588	0.0035
6	70	76.8118	-6.8118	46.4001
6	79	76.8118	2.1882	4.7884
4	64	64.0588	-0.0588	0.0035
5	75	70.4353	4.5647	20.8366
10	98	102.3176	-4.3176	18.6420
3	53	57.6824	-4.6824	21.9244
7	85	83.1882	1.8118	3.2825
SSE = 195.9766				

Computing Residuals Explained using the Regression Model: Regression Sum of Squares (SSR)

We must also compute the amount of error that has been reduced or explained when using the regression model. The name isn't very intuitive but the regression sum of squares (SSR) refers to the amount of squared residuals that are reduced when using the regression model compared to the alternative model. Think of the SSR is the amount of improvement we get when using the regression model instead of the alternative model. In keeping, the greater the SSR, the better the predictions are.

Let's take a moment to consider how SSR, SSE, and SST are connected to each other. SSR and SSE are in opposition to one another. Recall that SSE refers to residuals left unexplained by the regression model. The lower the SSE, the better the predictions are. When SSR is higher, SSE is lower and when SSR is lower, SSE is higher. Recall also that the SST is the sum of the SSR and SSE. Therefore, we can find any one of these three if we already know the other two.

Here is a summary of the three forms of sum of squared residuals and how they are connected:

Error in the Alternative Model = Error Explained Using the Regression Model + Error Left Unexplained Using the Regression Model

SST = SSR + SSE

1,924 = SSR + 195.9766

SSR can, thus, be found by subtracting SSE from SST. For Data Set 12.1, these computations are as follows:

1,924 – 195.9766 = SSR

1,924 – 195.9766 = 1,728.0234

Thus, the SSR is easily and quickly computed using the SST and SSE.

Table 13.3. Symbols and Corresponding Formulas for Y

Symbol	Meaning	Formula
Y	Raw or observed score for a Y-variable	Given in a data set
$\bar{Y}$	The mean of scores for a Y-variable	$\bar{Y} = \frac{\Sigma Y}{n}$
$\hat{Y}$	A predicted value of Y	$\hat{Y} = b_1 x + b_0$
e	Residual; the difference between an observed value and its predicted value	$e = Y_i - \hat{Y}_i$

Table 13.4. Formulas for Computing Residuals

Symbol	Name in SPSS	Meaning	Formula	Steps
SST	Sum	The total squared error when using the alternative model	$\text{SST} = \Sigma(Y - \bar{Y})^2$	1. Fi

Symbol	Name in SPSS	Meaning	Formula	Steps
	of Squares Total		or SST = SSR + SSE	nd $\bar{Y}$ (the mean of the Y-values) 2. Find each residual by subtracting the $\bar{Y}$ from each known Y-

Symbol	Name in SPSS	Meaning	Formula	Steps
				value 3. Square the residuals 4. Sum the squared residuals to get the SST Alternately, it can be found as the sum of SSR and SSE

Symbol	Name in SPSS	Meaning	Formula	Steps when those two values are already known.
SSE	Sum of Squares Residuals	The total squared error when using the regression model. This can also be thought of as the amount of error that is not explained or reduced by using the regression model	$\text{SSE} = \Sigma(Y - \bar{Y})^2$ or $\text{SSE} = \Sigma(e)^2$	1. Find each predicted $Y(\hat{Y})$ 2. Find the residual (error) for each case by subtracting predi

Symbol	Name in SPSS	Meaning	Formula	Steps
				cted $Y(\hat{Y})$ from observed $\boldsymbol{Y}$. 3. Square the residuals. 4. Sum the squared residuals.
SSR	Sum of Squares Regression	The total squared error that is reduced or explained when using the regression model. This can also be thought of as the amount of improvement when using the regression model.	$\text{SSR} = \Sigma(\hat{Y} - \bar{Y})^2$ or SSR = SST - SSE	1. Find each predict

Symbol	Name in SPSS	Meaning	Formula	Steps
				ed $Y(\hat{Y})$ using $\hat{Y} = \iota$ 2. Find the mean of $Y(\hat{Y})$. 3. Find the residuals for the mean of $\boldsymbol{Y}$ by subtracting the mean of $Y(\hat{Y})$ fr

Symbol	Name in SPSS	Meaning	Formula	Steps
				om each predicted $Y(\hat{Y})$. 4. Square the residuals. 5. Sum the squared residuals. Alternately, it can be found as the difference between

Symbol	Name in SPSS	Meaning	Formula	Steps
				SST and SSE when those two values are already known.

Note: The subscript i is used to denote individual scores.

Testing Goodness-of-Fit with ANOVA

When a regression is significant, it means a substantial enough proportion of the variance is accounted for or reduced using the regression model compared to how much is left unaccounted for by the model. To test this, regression uses ANOVA. Recall from Chapter 10 that ANOVA computes sum of squares between to represent variation that is systematic between groups and sum of squares within which represents variation that is not systematic (and occurs within groups). This same concept is used for an ANOVA within regression but with different names. Specifically, when ANOVA is used in regression, the regression sum of squares (SSR) represents the systematic variation (i.e. residuals which are accounted for by the regression model) and the sum of squares error (SSE) represents non-systematic variation (i.e. residuals which are unaccounted for by the regression model). These are used to assess the goodness-of-fit of the regression model.

The four components used to find F for a regression ANOVA are as follows:

1. SSR
2. SSE
3. df_M
4. df_E

We learned how to find SSR and SSE in the prior section. For this section, therefore, we will only add in how to compute the remaining two of these four components: df_M and df_E.

Degrees of Freedom for Regression

The df_M refers to the degrees of freedom for the regression model. This is equal to the number of predictors being used. In a simple, bivariate regression df_M = 1 because only one predictor variable is being used (the X-variable). Thus, for Data Set 12.1, df_M = 1.

The df_E refers to the degrees of freedom error. This is equal to the sample size (n) minus the number of predictors being used plus 1 like so:

$$df_E = n - (\text{number of predictors} + 1)$$

In a simple, bivariate regression $df_E = n - 2$ because only one predictor variable is being used (the X-variable) and if we add 1 to the number of predictors we get 2. Thus, for Data Set 12.1, we compute:

$$\begin{gathered} df_E = n - (\text{ number of predictors } + 1) \\ df_E = 10 - (1 + 1) \\ df_E = 10 - 2 \\ df_E = 8 \end{gathered}$$

13.10: Computing F

The formula for the ANOVA F in regression is as follows:

$$F = \frac{RSS \div df_M}{SSE \div df_E}$$

The numerator is used to calculate the mean sum of squares for the regression model (MSSR). The denominator is used to calculate the mean sum of squares error (MSSE). The final F-value indicates the ratio of residuals reduced using the regression model relative the residuals remaining when using the regression model. The higher the F-value, the more the regression model improved predictions and, thus, the greater the chance of a significant result. The lower the F-value, the less the model improved predictions and, thus, the lesser the chance of a significant result.

For Data Set 12.1, the four parts needed to compute F are as follows:

1. SSR = 1,728.0234
2. SSE = 195.9766
3. df_M = 1
4. df_E = 8

These are plugged into the F-formula and used to solve for F as follows:

$$F = \frac{1,728.0234 \div 1}{195.9766 \div 8} = \frac{1,728.0234}{24.4971} = 70.5399$$

When rounded to the hundredths place, this result is:

$$F = 70.54$$

This is a very large result indicating that the regression model reduces a much greater proportion of error (i.e. residuals) than it leaves unexplained. Another way to say this is that the regression model provides much better predictions of Y than the alternative model.

When using hand-calculations to test a hypothesis, significance is assessed using a critical value (CV). We will see how to find the CV and use it to determine significance in a later section. For now, we will move on to reviewing the other analyses that are needed for a complete regression analysis.

Secondary Analyses for Regression: **t-Testing and Slopes**

When a regression ANOVA is significant, the slope of the regression line is computed and tested to see whether it is why the regression model significantly improved predictions. When the *t*-test for the slope is significant, it means that the slope significantly improved predictions of Y. When using a bivariate regression, it is redundant to check the significance of the slope because it is the only slope so it must be the one improving predictions. For this reason, if the ANOVA is significant, the *t*-test will also be significant in simple (bivariate) regression.

> **When must slope significance be checked?**
>
> When an advanced version of regression is used, there are multiple predictors and, thus, multiple slopes (one for each predictor variable). When those models are used, the *t*-tests function like post-hoc analyses in one-way ANOVA. Thus, when a regression model with multiple predictors is being tested and has a significant ANOVA result, *t*-tests are used to assess which slopes were significantly contributing to improving the predictions and which, if any, were not. However, the significance of the slope does not need to be checked in bivariate regression because there is only one predictor. Therefore, if the ANOVA is significant, the slope of that predictor is also significant and when the ANOVA is not significant, the slope of that predictor is also non-significant.

When the *t*-test is run using SPSS for Data Set 12.1, the results are shown by the software as follows:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	38.5529	5.177		7.447	<.001
	Hours of Sleep	6.3765	.759	.948	8.399	<.001
a. Dependent Variable: Quiz Scores						

The first row of the table shows the results for the y-intercept and the second row shows the results for the slope. Because the slope is the focus of the *t*-test, we will focus on the second row of results. It shows the name of the X-variable (Hours of Sleep), the slope of the line (6.3765), and the results of the *t*-test assessing the usefulness of that slope.

The obtained *t*-value for the slope, when rounded to the hundredths place, is 8.40. The "Sig." stands for "significance;" the value shown in that column is the *p*-value (which is the risk of a Type I Error). When $p < .05$, a result is significant. When $p \geq .05$, a result is *not* significant. The sig. value shown in the SPSS output is shown as "<.001" which means the risk of a Type I Error is less than 0.1%. This indicates that the slope of the regression line is significantly contributing to the improvement in predictions when using the regression model.

Interpreting Slopes

When a bivariate regression model is significant, the slope is often interpreted and reported as part of a complete results paragraph. The slope indicates the amount of change in the Y-variable that is predicted for every one unit increase in the X-variable. The slope, when rounded to the hundredths place, is 6.38. The X-variable is Hours of Sleep and the Y-Variable is Quiz Scores. Thus, the slope for Data Set 12.1 would be interpreted as follows:

> *For every one hour increase in sleep, there is a 6.38 unit increase predicted for quiz scores.*

Making Predictions

When a regression result is significant, the regression equation can be applied to make predictions. The formula used to make predictions was noted in an earlier section titled *Computing Residuals for the Regression Model: Sum of Squares Error (SSE)* so it will only be briefly reviewed here. The regression equation is written as follows:

$$\hat{Y} = b_0 + b_1 x$$

The necessary parts for this equation for Data Set 12.1 are as follows:

b_0: Y-intercept	38.5529
b_1: Slope of Hours of Sleep	6.3765
Note: Values continue but are shown to the fourth decimal place for space.	

Thus, the predicted Y-values for the regression model with Data Set 12.1 are computed using the following regression equation:

$$\hat{Y} = 38.5529 + 6.3765x$$

In the prior section, we used this equation to predict the Y-values using the X-values in Data Set 12.1. This was so we could compare the predictions to the actual Y-values in the data set. However, it can be used to predict using any X-value, not just those in the data set. To predict, take a given X-value, plug it into the equation, and compute $\hat{Y}$. Here is what it looks like to compute and interpret predicted Y given three different X-values:

Example 1	**Example 2**	**Example 3**

Example 1	**Example 2**	**Example 3**
Given X = 0.00 $\hat{Y}$ = 6.3765(0.00) + 38.5529 $\hat{Y}$ = 38.5529 The predicted quiz score when someone has gotten 0.00 hours of sleep is approximately 38.55 points.	Given X = 4.00 $\hat{Y}$ = 6.3765(4.00) + 38.5529 $\hat{Y}$ = 64.0589 The predicted quiz score when someone has gotten 0.00 hours of sleep is approximately 64.06 points.	Given X = 8.00 $\hat{Y}$ = 6.3765(8.00) + 38.5529 $\hat{Y}$ = 89.5649 The predicted quiz score when someone has gotten 0.00 hours of sleep is approximately 89.56 points.

Note: Slight error in being introduced by using rounded values for the slope and intercept.

Limitations

A major limitation of the regression, which it shares with correlation, is that it cannot be used to determine cause-effect relationships. Just because two things are mathematically related, does not mean that either is the cause of another. Therefore, though tempting, it is not generally appropriate to use causal language when interpreting the results of correlation nor regression (see Chapter 8 for a review of causal language). Some of the language can be misunderstood to be causal such as the use of the terms *predicted* and *explained*. However, these do not indicate that cause-effect has been determined. Instead, predictions are estimations of what is expected based on the current regression model. When we refer to the amount of variance that is "explained" we simply mean the amount that is accounted for based on the pattern among the data and corresponding regression line. When we say something is explained, we have not (and generally cannot) determine whether the relationship is causal and, if so, which variable is the cause. Therefore, it is important not to presume nor to indicate that a regression result is sufficient to determine that X caused Y.

1. What does SST represent?
2. How is SST calculated?
3. What does SSE represent?
4. How is SSE calculated?
5. What does SSR represent?
6. How is SSR calculated?
7. How is df_M calculated?
8. How is df_E calculated?
9. What is the formula for calculating F?

13.11: Example of How to Test a Hypothesis Using Regression

Let's walk through the steps for testing the hypothesis that sleep hours will be useful in predicting quiz scores using Data Set 12.1. This will include summarizing some of the computations we learned in prior sections of this chapter.

Steps in Hypothesis Testing

In order to test a hypothesis, we must follow these steps:

1. **State the hypothesis.**

A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that hours of sleep will be useful for predicting quiz scores.*

Non-Directional Hypothesis for a Regression

Research hypothesis	Hours of sleep will predict quiz scores.	$H_A : \beta_1 \neq 0$
Null hypothesis	Hours of sleep will not predict quiz scores.	$H_0 : \beta_1 = 0$

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The usefulness of one quantitative variable for predicting another quantitative variable is being tested so the appropriate test is simple regression.

3. **Determine the critical value.**

In order to determine the critical value for a regression, three things must be identified:

1. the alpha level,

2. the degrees of freedom for the model (df_M), and

3. the degrees of freedom for the error (df_E)

The alpha level is often set at .05 unless there is reason to adjust it such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The df_M and the df_E must also be calculated. The df_M is equal to the number of predictor variables. In simple bivariate regression only one predictor variable is used (the X-variable) so the $df_M = 1$. In simple bivariate regression the $df_M = n - 2$. Thus, the two forms of df for Data Set 12.1 are as follows:

$$df_M = 1$$

$$df_E = 8$$

The alpha level and dfs are used to determine the critical value for the test. The full tables of the critical values for F are located in Appendix E. Under the conditions of an alpha level of .05, $df_M = 1$, and $df_E = 8$, the critical value (CV) is 5.318 (see Appendix E).

CV = 5.318

The critical value represents the value which must be exceeded in order to declare a result significant. It represents the threshold of evidence needed to be confident a hypothesis is true. Regression uses ANOVA for which the result can only be a positive value. Thus, the obtained F-value must be greater than 5.318 to be declared significant when using Data Set 12.1.

4. **Calculate the test statistic.**

An ANOVA F-value is used to see if an X-variable is useful for predicting a Y-variable. The formula for this is:

$$F = \frac{RSS \div df_M}{SSE \div df_E}$$

See the section titled *Testing a Regression Model with ANOVA* earlier in this chapter for details on how SSR and SSE are calculated. For this section, we will show the abbreviated computations once those values and the dfs are known. For Data Set 12.1, the results are computed as follows:

$$F=\frac{1,728.0234 \div 1}{195.9766 \div 8}=\frac{1,728.0234}{24,4971}=70.5399$$

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value for F exceeds the critical value as follows:

The critical value is 5.318.

The obtained F-value, rounded to the hundredths place, is 70.54.

The obtained F-value exceeds (i.e. is greater than) the critical value, thus, the result is significant.

6. **Calculate the effect sizes and relevant secondary analyses.**

When it is determined that the result is significant, effect sizes should typically be computed. However, in correlation and regression, a secondary analysis is typically given rather than an effect size. Thus, the foci of step 6 for regression is to calculate and interpret the coefficient of determination and the slope of the regression line.

The **coefficient of determination** is the percent of variation in the Y-variable that is accounted for by variance in the X-variable. It can also be described as a calculation of how well a model using one variable (X) can be used to estimate the other (Y). These refer to two ways of interpreting and describing the same thing with the former more applicable to correlation and the latter more applicable to regression. The coefficient of determination results and interpretation will be summarized here as they were already covered for Data Set 12.1 in Chapter 12. For a full review of how to compute and interpret this value, see Chapter 12.

The symbol and formula for the coefficient of determination are the same and are written as follows:

$$r^2$$

The coefficient of determination for Data Set 12.1 would be computed as follows:

$$r^2=(0.9477\ldots)^2$$
$$r^2=0.8981\ldots$$

This result is quite large and can be reported as a percent and interpreted as follows for Data Set 12.1:

Approximately 89.8% of the variance in quiz scores was accounted for by variance in hours of sleep.

In addition, when a regression ANOVA is significant, the slope should be interpreted and supported with an evidence string for the *t*-test. These are generally computed using SPSS rather than by hand and are as follows:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	38.5529	5.177		7.447	<.001
	Hours of Sleep	6.3765	.759	.948	8.399	<.001

a. Dependent Variable: Quiz Scores

The slope, when interpreted and supported with the *t*-test results, would be reported as follows:

For every one hour increase in sleep, there is a 6.38 unit increase predicted in quiz scores, t(8) = 8.40, p < .05.

Note that the df_E is used for the *t*-test in a regression.

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

Following this, the results for our hypothesis with Data Set 12.1 can be written as shown in the summary example below.

APA Formatted Summary Example

A simple regression was used to test the hypothesis that hours of sleep would predict quiz scores. Consistent with the hypothesis, hours of sleep was a significant predictor of quiz scores, $F(1, 8) = 70.54$, p < .05. Approximately 89.8% of the variance in quiz scores was accounted for by variance in hours of sleep. For every one hour increase in sleep, there is a 6.38 unit increase predicted in quiz scores, t(8) = 8.40, p < .05.

As always, the APA-formatted summary provides a lot of detail in a particular order. To understand how to read and create a summary like this, review the detailed walk-through in Chapter 7. For a brief review of the structure for the APA-formatted summary of the omnibus test results, see the summary below.

The following breaks down what each part represents in the evidence string for the ANOVA and *t*-test results in the APA-formatted paragraph above:

Symbol for the test	Degrees of Freedom	Obtained Value	p-Value
F	(1, 8)	= 70.54	p < .05.
t	(8)	= 8.40	p < .05.

1. How is the ANOVA part of regression interpreted and reported in APA-format?
2. How is the *t*-test part of regression interpreted and reported in APA-format?
3. When a regression is significant, how is the coefficient of determination interpreted and reported?
4. When a regression is significant, how is the slope interpreted and reported?

13.12: Testing a Hypothesis with Regression in SPSS

This section focuses on how to analyze data for a simple regression using SPSS. SPSS version 29 was used for this book; if you are using a different version, you may see some variation from what is shown here.

Entering Data

Data for correlation and regression are set up the same way. See the section on entering data in Chapter 12 as needed. Here is how Data Set 12.1 looks after data are entered into SPSS:

Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

Conducting a Simple (Bivariate) Regression in SPSS

The steps to running a simple, bivariate regression in SPSS are:

1. Click Analyze -> Regression -> Linear from the pull down menus.
2. Drag the names of the predictor (X -variable) into the box that says "Independent(s)" and the predicted (Y -variable) into the box that says "Dependent." You can also do this by clicking on the variable names to highlight them and the clicking the arrow to move each of them to the desired location.

3. Click "OK" to run the analyses.
4. The output (which means the page of calculated results) will appear in a new window of SPSS known as an output viewer. The results will appear in three tables as shown below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.948[a]	.898	.885	4.94945
a. Predictors: (Constant), Hours of Sleep				

ANOVA[a]

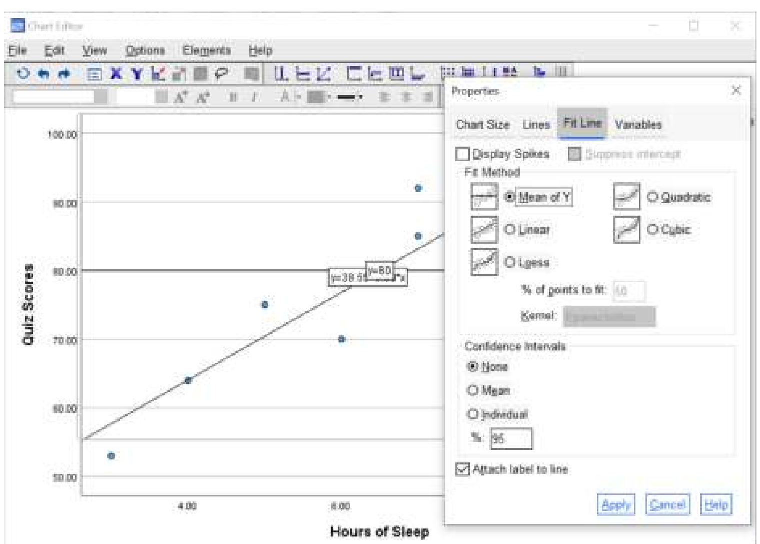
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1728.024	1	1728.024	70.540	<.001[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
	Residual	195.976	8	24.497		
	Total	1924.000	9			
a. Dependent Variable: Quiz Scores						
b. Predictors: (Constant), Hours of Sleep						

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	38.553	5.177		7.447	<.001
	Hours of Sleep	6.376	.759	.948	8.399	<.001
a. Dependent Variable: Quiz Scores						

Note: Slight error was introduced in hand-calculations due to rounding which can cause those values to differ from the SPSS results in the third to fourth decimal places.

5. If a scatterplot is desired to visualize the regression, click Graphs -> Scatter/Dot -> Simple Scatter -> Define. Move the names of X- and Y-variables from the left to their respective boxes on the right. Then click "OK." The graph will appear in the output window.
6. If a regression line is desired, double-click the graph to open the graph editor. Then click the Fit Line button. Choose the "Linear" option for the fit line and click "Apply." If the mean of Y line is also desired, click the Fit Line button again and then select "Mean of Y." You can also double click on an existing fit line to reopen the window to change which type it is (such as from a linear regression line to a mean of Y line).



Reading SPSS Output for a Simple Regression

The Model Summary Table shows the correlation coefficient which is r = .948. This is .95 when rounded to the hundredths place for reporting purposes. It also shows the r-squared value which is .898 or 89.8%. These match the results from the hand-calculations performed in Chapter 12 for Data Set 12.1.

The ANOVA table shows the parts that go into the F formula, the resulting F-value, and the "sig." The computed values match those which were hand-calculated earlier in this chapter to the hundredths place. Recall that "Sig." is SPSS refers to the p-value.

When this is less than the alpha level of .05, it is determined that the result is statistically significant and the hypothesis is supported. When this is greater than or equal to the alpha level of .05, it is determined that the result is not statistically significant and the null hypothesis is retained.

The coefficients table is used to find the slope and the y-intercepts as well as to assess whether the *t*-test was significant. The slope (b_1) and y-intercept (b_0) appear in the B column. The slope appears in the bottom row of the B column next to the name of the X-variable. The y-intercept appears on the top row of the B column next to the label "(constant)." These are the values that are put into the linear equation when X is used to predict a Y-value. The *t*-test results for testing the slope appear on the bottom row of the coefficients table. The last two columns present the obtained *t*-value and the corresponding p-value ("Sig."). Note that the table does not show the df for the *t*-test but that this is because the df_E is used for the *t*-test in a regression. Thus, the df for the *t*-test can be found in the ANOVA results table.

1. What information is provided in the Model Summary Table of SPSS output for simple regression?
2. What information is provided in the ANOVA Table of SPSS output for simple regression?
3. What information is provided in the Coefficients Table of SPSS output for simple regression?
4. Which kind of graph and fit lines should be used to visualize a simple regression?

CHAPTER OVERVIEW

14: Chi-Squared

14.1: Introduction to Non-Parametric Tests

All of the inferential tests we have covered so far in this book are parametric tests. Parametric tests use data for quantitative variables measured on interval or ratio scales. In *t*-tests and ANOVAs, we compared scores for quantitative variables between and among groups. In correlation and regression we assessed and used patterns between quantitative variables. For all of these tests, means and standard deviations can be computed and used to summarize the distribution of the quantitative variable(s). However, some hypotheses are simply focused on comparing counts for qualitative variables or quantitative variables measured on the ordinal scale and thus, do not include data which fit these parametric tests. When this occurs, non parametric test are needed.

Chi-squared tests are non-parametric tests used to compare the counts of subgroups for data measured on the nominal or ordinal scales of measurement. For this chapter, we will focus on two versions of chi-squared:

1. The Chi-Squared Goodness of Fit test and
2. Chi-Squared Test of Independence.

14.2: Chi-squared Goodness of Fit

There are times when a hypothesis makes a statement about whether counts or proportions of subgroups in a sample will be approximately equal to specified counts or proportions. In these situations, a chi-squared goodness of fit test is appropriate. **The chi-squared goodness of fit** tests whether the counts of subgroups of a variable fit specified, expected proportions or counts. One thing that makes the chi-squared quite unusual is that the null hypothesis indicates the specified counts and, thus, sometimes a null result is desired. This is because the null states that the counts observed in the data will match those specified when using a goodness of fit test and the alternative hypothesis would, thus, be that the counts in the data do *not* match those specified.

Usually the chi-squared goodness of fit is used to test whether the distribution of data from a sample is similar to the distribution that is known or presumed true about a population. Chi squared goodness of fit, therefore, is assessing whether the data are a good fit to (i.e. are similar or dissimilar to) the counts or proportions specified in the null hypothesis. When a result is significant, it means that the counts are significantly different than those specified by the null hypothesis. When a result is not significant, it means that the counts are approximately equal to those specified by the null.

Let's take a look at this using an example. Suppose that among the population of college students, 25% were business majors, 25% were psychology majors, 20% were nursing majors, and 30% were statistics majors. Suppose that you wanted to test whether the proportion of students in each of these majors at a small college in Statistonia students, overall. In this case, you would expect that at the proportions of majors at Statistonia College would be approximately 25% business majors, 25% psychology majors, 20% nursing majors, and 30% statistics majors. In this case, these percentages would be used as the specified proportions and a non-significant result would mean that the counts observed at Statistonia College were *not significantly different* from those in the population. A significant result would mean that the counts observed at Statistonia College *were significantly different* from those in the population. This is because a significant result occurs when counts observed in the data are different from the expected counts. (Note: We will see how to test this specific example in the SPSS section).

There are two kinds of counts considered in a chi-squared goodness of fit: Observed counts and expected counts. Observed counts refer to the counts that exist in the sample data. The expected counts are based on either known population proportions or on a hypothesis about what counts are expected to be. We will review these in detail later in this chapter. For now, let's focus on being clear on the general idea of this test. Below shows a summary of how to interpret results from a chi-squared goodness of fit. Refer back to this as needed to remind yourself what the test is comparing and what a significant result means compared to what a non-significant result means.

Interpreting Significance in a Chi-squared Goodness of Fit

Significant Result	Counts observed in the sample are significantly different from those expected based on the population or a hypothesis.	$Observed\ Counts \neq Expected\ Counts$
Non Significant Result	Counts observed in the sample are *not* significantly different from those expected based on the population or a hypothesis. *Alternative wording*: Counts observed in the sample are approximately equal to those expected based on the population or a hypothesis.	$Observed\ Counts \approx Expected\ Counts$

Variables

One or more grouping/categorical variables are used in chi-square. These variables can either include nominal or ordinal data which are separated into categories for the purposes of hypothesis testing. It is also possible to create categories with interval or ratio data using intervals, however, this is rarely used because other tests are generally a better fit to those data and their corresponding hypotheses.

In a univariate chi-square, one variable with at least two categories or levels is used. However, it is possible to use a version that looks at the counts of categories of two or more variables at once such as when testing whether the counts by gender and race/ethnicity in a sample are proportionally similar to those in the population. However, the focus of this section of the chapter will be univariate chi-square.

Data and Assumptions

Each statistical test has some assumptions which must be met in order for the formula to function properly. In keeping, there are a few assumptions about the data which must be met before a chi-squared goodness of fit is used. The most important is that no category can have an expected count lower than 5. Thus, sample sizes should be large enough that each expected count is 5 or higher.

Hypotheses

Hypotheses for chi-squared goodness of fit focus on whether observed counts are similar or significantly different from those expected based on populations or theories (and their corresponding hypotheses). We will focus on the non-directional hypothesis for this chapter because the goodness of fit test functions as an omnibus test. It can indicate whether counts are or are not as expected but cannot, on its own, indicate which counts are different than expected and which are not. When there are only two subgroups, only the chi-squared goodness of fit test is needed to know whether both groups were significantly different than expected. When there are three or more groups and a significant result, however, some ambiguity remains about the pairwise comparisons. Thus, what the goodness of fit test tells us is, overall, whether counts of subgroups are as expected or not. The non-directional research and corresponding null hypotheses for chi-squared goodness of fit can be summarized as follows:

Non-Directional Hypothesis for Chi-squared Goodness of Fit

Research hypothesis	Counts observed in the sample are significantly different from those expected based on the population or a theory.	$H_A: f_{\text{observed}} \neq f_{\text{expected}}$
Null hypothesis	Counts observed in the sample are *not* significantly different from those expected based on the population or a theory.	$H_0: f_{\text{observed}} = f_{\text{expected}}$

In this example \(f_{\text {observed }}\) refers to the frequency of counts observed in the data for each subgroup or category of the test variable and \(f_{\text {expected }}\) refers to the frequency of counts expected in the data for each subgroup or category of the test variable based on what is known or theorized about the population. When the counts in the sample are similar or equal to those expected, the result will be non-significant and, thus, favor the null hypothesis. Conversely, when the counts in the sample are notably dissimilar from those expected, the result will be significant and, thus, support the alternative hypothesis.

Limitations

A limitation of chi-squared goodness of fit is that it cannot always be used to determine cause effect relationships; however, under some conditions this may be appropriate using a different form of chi-squared known as a chi-squared test for independence (which we will cover later in this chapter). In addition, the chi-squared goodness of fit test can be hard to interpret when there are many categories and may not be an option when any of those groups has an expected count that is less than 5. However, it otherwise is a good option for comparing group counts when sufficient data are available.

Reading Review 14.1

1. What kinds of variables are tested using chi-squared goodness of fit tests?
2. What does a significant result indicate when using chi-squared goodness of fit tests?
3. What does a non-significant result indicate when using chi-squared goodness of fit tests?

14.3: Testing Goodness of Fit with Chi-Squared

Testing a Goodness of Fit Hypothesis

This test is used to see whether counts of subgroups or categories are approximately as expected (based on population data, underlying theory, and/or hypothesis) or are significantly different than expected. When the counts observed in the data are similar to the expected counts, the expected counts are said to be a good fit to the data. When the observed counts are significantly dissimilar to the expected counts, the expected counts are a poor fit to the data.

Determining Expected Counts

Expected counts are set based on either what is known or expected to be true about populations. Let's review the origins for each of these possible ways of setting expected counts. When a sample is being compared to a known population, the population statistics proportions or counts are used to set expectations. When this is done, chi-squared is often being used to see if a sample is an appropriate match to, or representative of, the population from which it was drawn. This can be very useful if you want to establish that a sample represents a population on key demographics such as age, gender, ethnicity, and/or socioeconomic status before using them to test other hypotheses about the population. If the sample is not significantly different from the population on any of the demographic variables, they may be said to be a good representation of the population.

In this same way, a chi-squared goodness of fit can be used to test whether a sample is significantly different from a population or another sample when this may be relevant. For example, it can be used to see whether the counts of different ethnic identities in a sample of students is similar to (or significantly different from) the proportions of those identities in the rest of the student body they are meant to represent. If the counts are similar, it may be determined that the sample is a good representation of the sample on this specific demographic. If the counts are significantly different from the population of students, however, it may be determined that the sample is biased and/or that generalizations should be limited.

A chi-squared goodness of fit can also be used to see if a change has occurred in population. For example, suppose that a company does quality control testing each year and in the prior year, 80% of the products passed the quality control test but 20% failed and needed to be discarded. Suppose that since then the company has made substantial changes to their process to try to reduce quality issues. Suppose that after these changes, they do a sample quality control test of 100 products and find that 96 of them (equivalent to 96% of the sample) passed the quality control test and 4 (equivalent to 4%) fail and need to be discarded. The company could test whether this is a significant improvement over what would be expected before they made those changes to their process. In this example, a significant result would be desirable as it would be taken to indicate that the changes to their process resulted in significantly fewer items needing to be discarded due to quality issues.

The other way expected counts can be set is based on something theorized, hypothesized, or otherwise desired. For example, suppose that a company is launching a new drink and want to know which of two versions to release. Based on the success of other companies, they believe a sweeter version of the drink will be preferred and sell better than a less sweet, healthier version. Suppose they wanted to know whether the number of customers who preferred the sweet version would be different than the number of customers who preferred the healthier version. A chi-squared goodness of fit could be used to test whether the counts of those who preferred each version were even, indicating no clear preference for either version, or significantly uneven, indicating a preference for one version over the other. Knowing whether there is a significant preference for one version over the other can be very valuable information for the company when making their decision about which version to release.

Computing Expected Counts from Percents

When expected counts are based on percentages, those percentages need to be converted to counts before they can be used in the goodness of fit formula. To convert from a percentage to an expected count, the following formula is used:

$$f_e = pn$$

Where:

f_e stands for frequency (i.e. count) expected for a category

p stands for the proportion or percent expected for a category in decimal form

n stands for the sample size for the data set

For example, suppose there was a sample of 40 adults, of which 40% were expected to be have children and 60% were expected to not have children. The expected counts for each subgroup would be calculated as follows:

Has Children (subgroup 1):

$$f_{e1} = .40(40) = 16$$

Does Not Have Children (subgroup 2):

$$f_{e2} = .60(40) = 24$$

Thus, the expected count for those who had children would be 16 persons and the expected count for those who did not have children would be 24 persons. Note that when you sum the expected counts for all groups, it will be equal to the total sample size.

Determining Observed Counts

Observed counts are based on sample data. Each category for a variable is identified and the participants in each of those categories are simply counted to get the observed counts. These are called observed counts because data are examples of what has been observed. The symbol used for frequency observed is *fo*. Numbers or names can be added to the subscript to differentiation subgroups. For example, if there were two subgroups, f_{o1} could be used to refer to the observed counts for subgroup 1 and f_{o2} could be used to refer to the observed counts of subgroup 2.

The Goodness of Fit Formula

The chi-squared goodness of fit formula is fairly simple and is as follows:

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

The steps to using this formula are as follows:

1. Find the difference between f_o (frequency observed in the data) and f_e (the frequency expected) for each category (i.e subgroup).
2. Square the difference for each category.
3. Divide the squared difference by f_e for each category.
4. Sum the results of step 3 to get the χ^2 value.

Thus, the $\frac{(f_o - f_e)^2}{f_e}$ portion must be computed for each category or subgroup before these values are summed. If there are two categories, $\frac{(f_o - f_e)^2}{f_e}$ is computed twice, once for each category, before their results are summed. If there are three categories, $\frac{(f_o - f_e)^2}{f_e}$ is computed three times, once for each category, before their results are summed, and so on.

Example Using the Goodness of Fit Formula

Let’s try using the formula to test whether the counts of sample members who did and did not have children was approximately equal to 40% and 60%, respectively. Suppose that in the sample of 40 adults, 22 did not have children and 18 did. These would be used for the observed counts. The data can be organized as follows:

Counts for Goodness of Fit Test

Subgroups	Observed	Expected
Has Children	18	16
Does Not Have Children	22	24

There are two ways you can organize and compute the χ^2 using the goodness of fit formula: table version and formula version. I really like the table format for organizing and conducting computations but some find the formula version easier to understand. The computations and steps are the same in both versions. The only difference is how the information is laid out. Thus, both versions are equally useful and are shown here so you can choose which the format that is the clearest to you.

Table format builds from the summary counts table above to organize and complete the chi squared formula steps. The observed and expected counts are the preparatory work which are then used in the steps of the chi-squared goodness of fit formula. For the above data, the table format of steps and results are as follows:

Table 14.1 Computations for Goodness of Fit in Table Format

Preparation			Steps		
Subgroups	**Observed**	**Expected**	**Differences** $f_o - f_e$	**Squared** $(f_o - f_e)^2$	**Divided** $\frac{(f_o - f_e)^2}{f_e}$
Has Children	18	16	2	4	4/16 = 0.25
Does Not Have Children	22	24	(-2)	4	4/24 = 0.1667
Total	Summed 40				Summed $\chi^2 = 0.4167$

The result when rounded to the hundredths place is: $\chi^2 = 0.42$.

The same preparation and steps can be used in formula format. For this format, the parts of the formula are filled in and computed using order of operations. Here is what it looks like to compute the χ^2 in formula format:

Table 14.2 Computations for Goodness of Fit in Formula Format

		Has Children		Does Not Have Children
χ^2	=	$\frac{(f_o - f_e)^2}{f_e}$	+	$\frac{(f_o - f_e)^2}{f_e}$
χ^2	=	$\frac{(18-16)^2}{16}$	+	$\frac{(22-24)^2}{24}$
χ^2	=	$\frac{(2)^2}{16}$	+	$\frac{(-2)^2}{24}$
χ^2	=	$\frac{4}{16}$	+	$\frac{4}{24}$
χ^2	=	0.25	+	0.1667
χ^2	=		0.4167	

The result when rounded to the hundredths place is: $\chi^2 = 0.42$.

Notice that the steps and results are the same whether you use the table format or the formula format. This is because each is just a different way of showing the same steps and computations.

Reading Review 14.2

1. What is an observed count based on?
2. What is an expected count based on?
3. What does f_o represent?
4. What does f_e represent?
5. What are the steps to computing χ^2 goodness of fit for two categories?

14.4: Example of How to Test a Hypothesis Using Chi-squared Goodness of Fit

Suppose a café is preparing to offer four different flavors of latte for the fall season but they think the preferences of customers for the four flavors will be uneven. They do a test run wherein they have 120 participants try each latte and select their favorite. Let's test the hypothesis that the counts of customers who prefer each of the four different flavors will be uneven using Data Set 14.1. We will use this information to follow the steps in hypothesis testing.

Data Set 14.1. Preferred Drink Flavor.

Counts for Goodness of Fit Test	
Flavors	Observed Counts
No Flavor Added	30
Vanilla	39
Chocolate	36
Pumpkin	15

Steps in Hypothesis Testing

In order to test a hypothesis, we must follow these steps:

1. **State the hypothesis.**

A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that the counts of preferences will be uneven across different drink flavors.*

Non-Directional Hypothesis for Chi-squared Goodness of Fit

Research hypothesis	Counts of preference for different drink flavors will be significantly uneven.	$H_A : f_{\text{observed}} \neq f_{\text{expected}}$
Null hypothesis	Counts of preference for different drink flavors will *not* be significantly uneven.	$H_0 : f_{\text{observed}} = f_{\text{expected}}$

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The counts of categories for a qualitative variable are being tested so the appropriate test is chi-squared goodness of fit.

3. **Determine the critical value.**

In order to determine the critical value for chi-square, we need to know the alpha level and the degrees of freedom. The alpha level is often set at .05 unless there is reason to adjust it such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The degrees of freedom for chi-squared goodness of fit are computed using the following formula:

$$df = k - 1$$

Where k stands for the number of categories. In the current hypothesis there are 4 drink flavors so k = 4. Thus, the calculation for df for this example is as follows:

$$df = 4 - 1$$
$$df = 3$$

The alpha level and df are used to determine the critical value for the test. Below is the χ^2 critical values tables that fits the current hypothesis and data. Under the conditions of an alpha level of .05 and df = 3, the critical value is 7.815.

Chi-Squared Critical Values

df	α = .05	α = .01
1	3.841	6.635
2	5.991	9.210
3	7.815	11.345
4	9.488	13.277
5	11.070	15.086
6	12.592	16.812
7	14.067	18.475

The critical value represents the value which must be exceeded in order to declare a result significant. It represents the threshold of evidence needed to be confident a hypothesis is true. The obtained χ^2-value must be greater than 7.815 to be declared significant when using Data Set 14.1.

4. **Calculate the test statistic.**

In order to use a goodness of fit test, we first must find the observed and expected counts. Observed counts are based on a data set, however, expected counts must be computed. To find the expected counts, we need to know what the hypothesized proportions are. In the hypothesis, it states that the counts of the four groups will be *uneven*. This means that the null will state the counts *are even*. There are four categories being compared; if they had even counts (as stated by the null), it would mean that 25% of the sample would be in each of the four categories (because 25% is one-fourth). Now, we must find the total sample size and multiply it by 25% (which is 0.25 when written in decimal form) to find the expected counts.

Counts for Goodness of Fit Test

Flavors	Observed Counts	Expected Counts
No Flavor Added	30	.25 × 120 = 30
Vanilla	39	.25 × 120 = 30
Chocolate	36	.25 × 120 = 30
Pumpkin	15	.25 × 120 = 30
Total	120	

Notice that all the expected counts are the same. This will occur anytime we are testing whether counts are even or not. Now that we have the observed and expected counts for all categories, we can plug these values into the formula and solve. The computations for this example, shown in formula format, are as follows:

		No Flavor		Vanilla		Chocolate		Pumpkin
χ^2	=	$\frac{(f_o-f_e)^2}{f_e}$	+	$\frac{(f_o-f_e)^2}{f_e}$	+	$\frac{(f_o-f_e)^2}{f_e}$	+	$\frac{(f_o-f_e)^2}{f_e}$
χ^2	=	$\frac{(30-30)^2}{30}$	+	$\frac{(39-30)^2}{30}$	+	$\frac{(36-30)^2}{30}$	+	$\frac{(15-30)^2}{30}$
χ^2	=	$\frac{(0)^2}{30}$	+	$\frac{(9)^2}{30}$	+	$\frac{(6)^2}{30}$	+	$\frac{(-15)^2}{30}$
χ^2	=	$\frac{0}{30}$	+	$\frac{81}{30}$	+	$\frac{36}{30}$	+	$\frac{225}{30}$
χ^2	=	0.00	+	2.70	+	1.20	+	7.50
χ^2	=	11.40						

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value for χ^2 exceeds the critical value as follows:

The critical value is 7.815

The obtained χ^2-value is 11.40

The obtained χ^2-value exceeds (i.e. is greater than) the critical value, thus, the result is significant.

Keep in mind that obtained values are often rounded to the hundredths place when reported.

6. **Calculate the effect sizes and any secondary analyses.**

The chi-squared goodness of fit test is an omnibus test which can tell us whether, overall, the observed counts are different from the expected counts; it does not, however, always allow us to determine which category counts are different from their expected counts and which are not when not all counts are different than expected. Thus, post-hoc tests are sometimes desired when a chi-squared goodness of fit result is significant. When two categories are being compared, a post-hoc test is not generally used. However, when three or more category counts are being compared, various post-hoc tests (such as using a chi-squared test for independence with a Bonferroni correction to compare each pair of categories) may be desired and used. However, as the focus of this chapter is the omnibus test, we will not go into a detailed review of these secondary analyses. For our purposes, therefore, secondary analyses will not be used to test the hypothesis and the use of both the chi-squared goodness of fit test and the chi-squared test for independence will only be reviewed for use as omnibus tests.

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

There are no means or standard deviations to report for chi-squared because it is non parametric. It is, however, necessary to include the observed counts for each category in the results. Finally, it is customary to reported the total sample size using "N =" to the right of the df in the parenthesis of the evidence string. Following this, the results for our hypothesis with Data Set 14.1 can be written as shown in the summary example below.

APA Formatted Summary Example

A chi-squared goodness of fit was used to test the hypothesis that the counts of preference for different drink flavors would be significantly uneven. Consistent with the hypothesis, the counts of preference for non-flavored (n = 30), vanilla (n = 39), chocolate (n = 36), and pumpkin (n = 15) were significantly uneven, $\chi^2(3, N=120)=11.40$, p < .05.

As always, the APA-formatted summary provides a lot of detail in a particular order. For a brief review of the structure for the APA-formatted summary of the omnibus test results, see the summary below.

Anatomy of the Evidence String

The following breaks down what each part represents in the evidence string for the chi-squared results in the APA-formatted paragraph above:

Symbol for the test	Degrees of Freedom and Total Sample Size	Obtained Value	p-Value
χ^2	$(3, N=120)$	= 11.40	p < .05.

Reading Review 14.3

1. How is df calculated for a chi-squared goodness of fit test?
2. What is reported within the parenthesis next to df in the evidence string for chi-squared?

3. How are expected counts for each category calculated when testing whether counts are even or uneven using a chi-squared goodness of fit test?
4. What detail about each category should be included in the APA-formatted summary for a chi-squared goodness of fit test?

14.5: Testing Independence with Chi-Squared

Testing a Hypothesis of Independence

The chi-squared test for independence is also known as a Pearson's chi-squared test. This version of chi-squared is used to compare the counts of outcomes under different conditions and can be appropriate for experiments which are used to compare counts. Thus, cause-effect can sometimes be deduced using this technique when appropriate to the data and corresponding hypothesis. The formula for this is the same as the one for a goodness of fit. Where these techniques diverge is in how the expected counts are computed. Thus, the focus of this section will be the computations for the chi-squared test of independence first followed by completing the steps for hypothesis testing using an example.

A chi-squared test of independence can be used to test whether the counts for one variable are dependent on another variable. The test compares the observed frequencies to those that would be expected if the variables were independent. When the observed counts are similar to the expected counts and, thus, the result is non-significant, it indicates that the variables are independent. When the observed counts are significantly dissimilar to the expected counts, it indicates that the variables are dependent.

Suppose that you hypothesize that the count of customers who make a purchase depends upon whether they are greeted when entering a store. Suppose that to test this, every other customer is greeted when they enter the store until 80 customers have been observed. The summary data for the observations can be organized into a cross tabulation (crosstabs) table.

Data Set 14.2. Purchases by Greeting Condition.

Counts for Test of Independence			
Purchased	Did Not Purchase	Row	Total
Greeted	15	25	40
Not Greeted	8	32	40
Column Total	23	57	$N=80$

Determining Expected Counts

The expected counts are computed for each category using the following formula:

$$f_e = \frac{\text{row total } \times \text{ column total}}{N}$$

The expected counts for Data Set 14.2 for this scenario are as follows:

	Observed	Expected
Greeted		
Purchased	15	$f_e = \frac{40 \times 23}{80} = 11.50$
Did Not Purchase	25	$f_e = \frac{40 \times 57}{80} = 28.50$
Not Greeted		
Purchased	8	$f_e = \frac{40 \times 23}{80} = 11.50$
Did Not Purchase	32	$f_e = \frac{40 \times 57}{80} = 28.50$

The Chi-Squared Formula

The chi-squared formula is the same for the tests of independence and goodness of fit. Thus, the formula for computing χ^2 is still:

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

Recall that the steps to using this formula are as follows:

1. Find the difference between f_o (frequency observed in the data) and f_e (the frequency expected) for each category.
2. Square the difference for each category.
3. Divide the squared difference by f_e for each category.
4. Sum the results of step 3 to get the χ^2 value.

Example Using Chi-Squared Formula

Let’s complete the computations for Data Set 14.2 using the table method. The computations are as follows:

Table 14.2 Computations for Goodness of Fit in Table Format

Preparation			Steps		
Subgroups	**Observed**	**Expected**	**Differences** $f_o - f_e$	**Squared** $(f_o - f_e)^2$	**Divided** $\frac{(f_o - f_e)^2}{f_e}$
Greeted					
Purchased	15	11.50	3.50	12.25	12.25/11.50 = 1.0652…
Did Not Purchase	25	28.30	(-3.50)	12.25	12.25/28.50 = 0.4298…
Not Greeted					
Purchased	8	11.50	(-3.50)	12.25	12.25/11.50 = 1.0652…
Did Not Purchase	32	28.50	3.50	12.25	12.25/28.50 = 0.4298…
Total	80				$\chi^2 = 2.9900\ldots$

Steps in Hypothesis Testing

The computations for Data Set 14.2 are already shown above. Therefore, for this section, we will focus on the steps to testing but with an abbreviated section on the aforementioned computations. In order to test a hypothesis, we must follow these steps:

1. **State the hypothesis.**

A summary of the research hypothesis can be stated as follows: It is hypothesized that the count of customers who make a purchase depends upon whether they are greeted when entering a store.

The null hypothesis for this example would state that the counts of customers who make a purchase does not depend on whether they are greeted when entering a store. Keep in mind that the expected counts are what will occur if counts of purchases are *not* dependent on greetings. If the result is significant it will support the research hypothesis. However, if the result is not significant, it will not support the research hypothesis and the null will be retained.

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The counts of categories for a qualitative variable are being tested to see whether they are independent of another variable so the appropriate test is chi-squared test of independence.

3. **Determine the critical value.**

In order to determine the critical value for chi-square, we need to know the alpha level and the degrees of freedom. The alpha level is often set at .05. The degrees of freedom for this chi-squared are as follows:

$$df = k - 1$$
$$df = 4 - 1$$
$$df = 3$$

The alpha level and *df* are used to determine the critical value for the test. The tables of the critical values for χ2are located earlier in this chapter.

The critical value for this example is 7.815. The obtained χ^2-value must be greater than 7.815 to be declared significant when using Data Set 14.2.

4. **Calculate the test statistic.**

In order to use a test of independence, we first must find the observed and expected counts. Observed counts are based on a data set, however, expected counts must be computed. Computing the expected counts is the key feature which distinguishes a test of independence from a goodness of fit test when using chi-squared. To find the expected counts, the following formula is used:

$$f_e = \frac{\text{row total } \times \text{ column total}}{N}$$

The details to using this formula to compute the expected counts for Data Set 14.2 is shown in the prior section. Once found, the observed and expected counts are plugged into the same formula which was used for the goodness of fit tests. To remind, that formula is as follows:

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

The computations for Data Set 14.2 are shown in Table 14.2 in the previous section of this chapter (see Table 14.2 for step-by-step computations). The result, when rounded to the hundredths place, is: χ^2 = 2.99.

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value exceeds the critical value as follows:

The critical value is 7.815

The obtained χ^2-value is 2.99

The obtained χ^2-value does not exceed (i.e. is lesser than) the critical value and, thus, the result is *not* significant.

6. **Calculate the effect sizes and any secondary analyses.**

When a chi-squared test of independence result is significant, post-hoc tests are sometimes desired. When three or more category counts are being compared, various post-hoc tests (such as using a secondary chi-squared test for independence with a Bonferroni correction to compare each pair of categories) may be desired and used. However, because the current test was not significant, post-hoc analyses are not warranted.

7. **Report the results in American Psychological Associate (APA) format.**

The same formatting guidelines for reporting a goodness of fit result apply to a test of independence. In addition, in this example the expected counts were not even and this needs to be made clear. Thus, it can be useful to include the expected counts in the summary. The results for our hypothesis with Data Set 14.2 can be written as shown in the summary example below.

APA Formatted Summary Example

A chi-squared test of independence was used to test the hypothesis that the count of customers who make a purchase depends upon whether they are greeted when entering a store. Contrary to the hypothesis, the counts of those who were greeted who did (n = 15), and did not make purchases (n = 25) and the counts of those who were not greeted who did (n = 8) and did not make purchases (n = 32) were not significantly different than expected counts of 11.50, 28.50, 11.50, and 28.50, respectively, $\chi^2(3, N = 80) = 2.99$, p >.05. Thus, the data do not support the hypothesis that purchasing depends upon being greeted.

Reading Review 14.4

1. How is df calculated for a chi-squared test of independence?
2. How are expected counts for each category calculated when using a chi-squared test of independence?
3. What detail about each category should be included in the APA-formatted summary for a chi-squared test of independence?

14.6: Using SPSS

Let's return to our example about Statistonia College from the beginning of our chapter. For that example, we supposed that among the population of college students, 25% were business majors, 25% were psychology majors, 20% were nursing majors, and 30% were statistics majors. Suppose that you wanted to test whether the proportion of students in each of these majors at a small college in Statistonia were similar to that of college students, overall, in SPSS. In order to test a hypothesis using SPSS, we need the raw data rather than the summarized counts like we saw in Data Set 14.1. Thus, we will use Data Set 14.3 to test the current hypothesis using SPSS. Data Set 14.3 shows the qualitative data for majors for a sample of 30 in the left column. To help us summarize the hypothesized counts, the expected proportions for each category are shown in the right column but are *not* data.

Data Set 14.3. Majors at Statistonia College (n = 30)

Majors Data	Expected Proportions
Business	25%
Business	
Business	
Business	
Business	
Business	
Psychology	25%
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Nursing	20%
Nursing	
Nursing	
Nursing	
Nursing	
Nursing	
Nursing	
Statistics	30%
Statistics	
Statistics	
Statistics	
Statistics	

Majors Data	Expected Proportions
Statistics	

Data need to be organized and entered into SPSS in ways that serve the analysis to be conducted. Thus, this section focuses on how to enter and analyze data for a chi-squared goodness of fit using SPSS. SPSS version 29 was used for this book; if you are using a different version, you may see some variation from what is shown here.

Entering Data

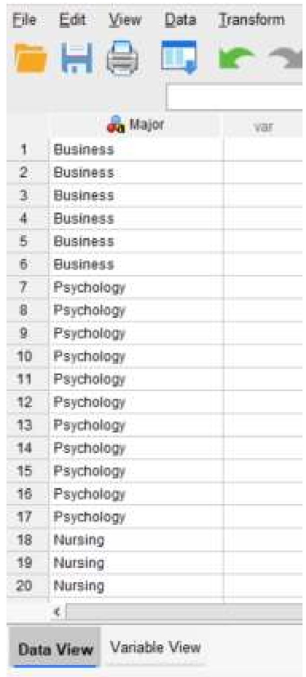
Open the SPSS software, click “New Dataset,” then click “Open” (or “OK” depending on which is shown in the version of the software you are using). This will create a new blank spreadsheet into which you can enter data. Click on the Variable View tab on the bottom of the spreadsheet. This tab of the spreadsheet has several columns to organize information about the variables. The first column is titled “Name.” Start here and follow these steps:

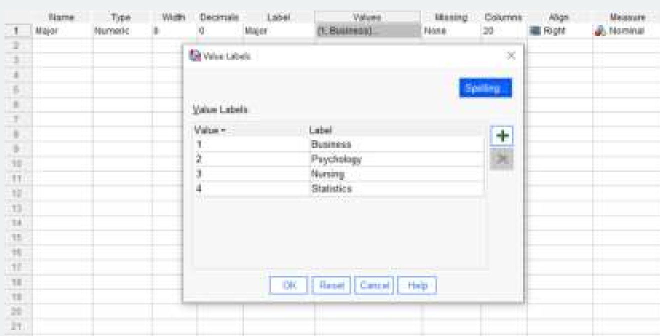
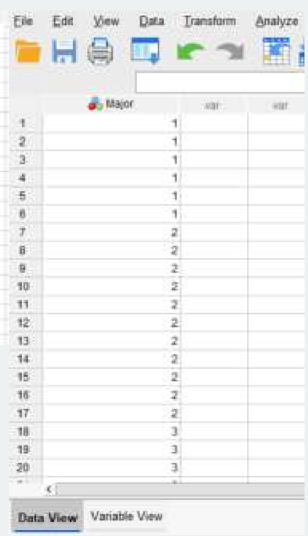
1. Click the first cell of that column and enter the name of your test variable using no spaces, special characters, or symbols. For Data Set 14.3 the variable name is Major. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
2. Click the first cell of the column titled “Type” and then click the three dots that appear in the right side of the cell. Ideally, we would select “string” which allows us to use words to name each group. However, SPSS works best if we recode each group name as number (i.e. 1 = Business, 2 = Psychology, 3 = Nursing, and 4 = Statistics). We will use this coding to enter our data. For type, specify that the data for that variable will appear as numbers by selecting “Numeric.” We will need to use numbers to represent each major.
3. Click the first cell of the column titled “Label.” This is where you can specify what you want the variable to be called in output, including in tables and graphs. You can use spaces or phrases here, as desired.
4. Click on the first cell of the column titled “Measure.” A pulldown menu with three options will allow you to specify the scale of measurement for the variable. Select the “Nominal” option because the variables for a chi-squared are qualitative.

Now you are ready to enter your data. Click on the Data View tab toward the bottom of the spreadsheet. This tab of the spreadsheet has columns into which you can enter the data. Click the cell of the first column and follow these steps:

1. Enter the data for the test variable moving down the rows under the first column. If your data are already on your computer in a spreadsheet format such as excel, you can copy-paste the data in for the variable.
2. Then hit save to ensure your data set will be available for you in the future.

Here is how the first 20 data points for Data Set 14.3 looks after data are entered into SPSS using the group names:

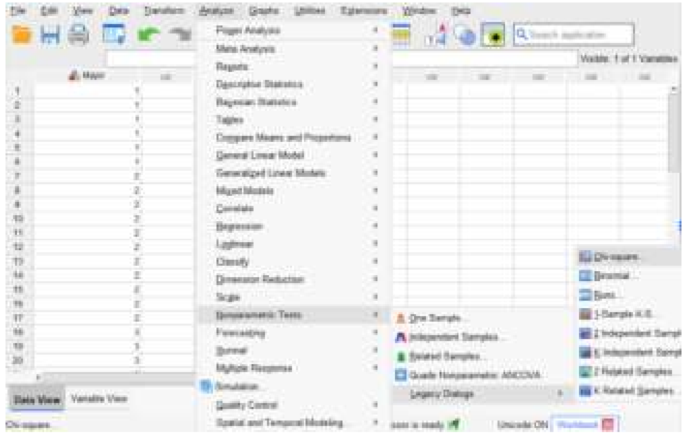


> **Note**
>
> SPSS Version 29 is fussy and only seems to work if you code the names with numeric values. Thus, if you get an error warning when trying to run your analysis, recode your majors as 1, 2, 3, and 4 in the data view tab and clarify what they stand for using the values cell of the variables tab as shown below:
>
> 
>
>
> 
>
>
> Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

Conducting a Chi-Squared Goodness of Fit Test in SPSS

The steps to running this form of chi-squared in SPSS are:

1. Click Analyze -> Nonparametric Tests -> Legacy Dialogs -> Chi-squared from the pull down menus.



2. Drag the name of the test variable from the list on the left into the box on the right of the command window. For Data Set 14.3, this means we are moving Major to the test box as shown below.

> **Note**
>
> When using SPSS Version 29, it may call your nominal variable ordinal at this step. This will not impact the analysis.

In this command window, we must set our expected values. We are testing whether the count of students who identified their majors as Business, Psychology, Nursing, and Statistics are similar to dissimilar to the proportions of 25%, 25%, 20%, and 30%, respectively. Thus, we need to enter those values as shown in the command window below:

3. Click “OK” to run the analyses.
4. The output (which means the page of calculated results) will appear in a new window of SPSS known as an output viewer. The results will appear in two tables as shown below:

	Major		
	Observed N	Expected N	Residual
Business	6	7.5	-1.5
Psychology	11	7.5	3.5

	Major		
	Observed N	Expected N	Residual
Nursing	7	6.0	1.0
Statistics	6	9.0	-3.0
Total	30		

Test Statistics

	Major
Chi-Square	3.100[a]
df	3
Asymp. Sig.	.376
a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 6.0.	

The first table summarizes the observed and expected counts for each category (i.e. for each major in this example). Notice that the expected percents have been transformed into their proportional expected counts. The second table provides the results of the inferential test using the observed and expected counts from the first table. The chi-squared formula result appears on the top row; the result is χ^2 = 3.10 when rounded to the hundredths place. The degrees of freedom appear in the second row. There were 4 categories whose counts were being compared and, thus, the df = 3. Finally, the p-value (which represents the risk of a Type I Error) appears in the third row as the “Asymp. Sig.”. This is the same as what SPSS has called “sig.” in our earlier chapters. When the p-value is less than .05, the result is significant and when it is greater than .05 the result is not significant. We see that the p-value for the present analysis was .376 and, thus, that the result is not significant.

Note

An assumption of a chi-squared test is that no category has an expected frequency lower than 5. SPSS automatically checks this assumption and alerts the user when it has been met or violated. We can see that SPSS has confirmed that this assumption was met for the present data in a note it has included under the second table of results.

The results from this test are as follows: A chi-squared goodness of fit test was used to test whether the counts of students in each major at Statistonia College were similar to those of students in the population, overall. Specifically, students overall majored in Business (25.0%), Psychology (25.0%), Nursing (20.0%), and Statistics (30.0%). The counts of students at Statistonia College who were majoring in Business (n = 6), Psychology (n = 11), Nursing (n = 7), and Statistics (n = 6) were not significantly different than the proportions in the population, $\chi^2(3, N=30)=3.10$, p >.05. Thus, the counts of each major at the college were proportionally similar to those in the population of students.

Reading Review 14.5

1. Where do you specify the expected proportions when setting up a chi-squared goodness of fit analysis is SPSS?
2. What information is provided in each of the two tables of SPSS output for chi-square?
3. Which assumption of chi-squared is automatically checked and reported by SPSS?
4. What does a non-significant result indicate when using a chi-squared goodness of fit test?

14.7: Structured Summary for Chi-Squared

After carefully reading the chapter, complete the following structured summary to add a learning check and easy-to-use reference to your notes.

Summarize what each symbol stands for.

N =

k =

f_o =

f_e =

Fill-in the appropriate information for each section below:

1. Chi-Squared Basics
   a. For which kinds of data can/should this be used?
   b. What is the focus of this statistic?
   c. What assumptions must the data meet to use this test?
2. Chi-Squared Formula
   a. What is the formula for chi-squared?
   b. What things should be computed in the preparatory steps for using this formula?
   c. What are the steps for solving using this formula?
3. Reporting Results from Chi-Squared.
   a. How is this statistic reported when using APA format?
      i. What things must be reported in the APA summary sentences for the chi squared?
      ii. When might it be useful or necessary to report the expected counts for a chi-squared in the APA-formatted results?

CHAPTER OVERVIEW

15: Appendix

15.1: Appendix A- Math Symbols and Their Operations

Below is a list of several commonly used symbols in math and statistics. Review the symbols, their meanings, and how they are used. The result of adding is called a *sum*. The result of subtraction is called a *difference*. The result of multiplication is called a *product*. The result of division is called a *quotient*.

Symbol	Interpretation	Example
=	is equal to	**10 = 10**
≠	is not equal to	**10 ≠ 5**
>	is greater than	**10 > x**
<	is less than	**x < 10**
≥	is greater than or equal to	**10 ≥ x**
≤	is less than or equal to	**x ≤ 10**
+	add/sum	**10 + 5 = 15**
−	subtract	**10 − 5 = 5**
× or ● or *	multiply	**10 × 5 = 50** **10 ● 5 = 50** **10*5 = 50**
÷ or / or —	divide	**10 ÷ 5 = 2** **10 / 5 = 2** $\frac{10}{5} = 2$
a^2 **or** *a*^2	square the number (i.e. multiply it by itself)	**if a = 5** $5^2 = 25$ **5^2 = 25**
$\sqrt{a}$	find the square root of the number	**if a = 25** $\sqrt{25} = 5$
[ ] or ( )	indicates expressions within the brackets or parentheses should be done first	**[(2+2) × (10-5)] = 20**
∑x	Add all instances of the variable X	**When X = 3, 4, and 5** **∑x = 12**

Note

When two sets of parentheses appear next to each other without any symbols between them, it indicates that the numbers within the parentheses are to be multiplied.

Here are some examples:

(8)(2) = 16 (3)(12) = 36 (2)(7) = 14 (20)(5) = 100 (4)(5) = 20

Evaluate the expressions.

15 × 5 =	$\sqrt{100}$ =	**10 - 5 =**
4^2 =	**100 × 8 =**	2^2 =
(5 - 2) =	$(-9)^2$ =	$\sqrt{9}$
(12 ÷ 2) =	**(4)(3) =**	10^2 =

3 × 15 =	$\sqrt{64}$	**75 ÷ 25 =**
12^2 =	**25 × 10 =**	**(5)(11) =**
6^2 =	**1 + 5 =**	$\frac{8}{1}$ =
$\sqrt{144}$ =	**81 ÷ 9 =**	**36 / 3 =**
14 ÷ 1 =	**40 / 2 =**	**17 × 0 =**

Fill in missing pieces using one of these symbols: = > <

The first problem has been completed for you.

15 > 11	**8 ____ -8**	**-5 ____ 0**
|-44| ____ 44	**4 ____ -15**	$\frac{1}{2}$ **____ 2**
9 ____ 88	**|7| ____ -2**	**-10 ____ 9**
-14 ____ -37	**-3 ____ 12**	**53 ____ 53**
43 ____ 4.3	**-1 ____ -0.1**	**-16.45 ____ 6.9**
-0.98 ____ -1.00	**0.25 ____ -2.50**	**-1,152 ____ 758**

15.2: Appendix B- Order of Operations

Some formulas require many operators (i.e. symbols that denote a mathematic procedure is to be performed). Mathematicians, statisticians, and scientists use agreed upon rules that govern how they should proceed in their work. These explicit rules reduce confusion and miscommunication among members of the fields. One such important rule is referred to as the Order of Operations. *The Order of Operations* refers to the order in which parts (i.e. steps) of equations must be carried out. The acronym PEMDAS is often used as a short-hand for the steps in their proper order: Parenthesis, Exponents, Multiplication and/or Division and/or, and finally Adding and/or Subtraction. Not following this order can cause an individual to yield a wrong answer.

When following the Order of Operations, you should conduct four steps:

1. First, complete all operations within parenthesis or brackets.
2. Second, calculate all the exponents.
3. Third, complete all multiplications and/or divisions in order from left to right.
4. Fourth, complete all addition and/or subtraction in order from left to right.

Note

Sometimes there are multiple steps within parentheses. When this is the case, you perform The Order of Operations **within** the parenthesis as step one.

This section will walk you through a few examples using The Order of Operations, starting with simpler equations and progressing to more complex equations.

Example 15.2.1

$70 - 6 \times 5$

Solution

There are only two operations to do: subtract and multiply. The Order of Operations tells us the complete all multiplication before subtraction. Therefore, the solution is 40 and is found as follows:

$70 - 6 \times 5 = 70 - 30 = 40$

Example 15.2.2

$4 \times 5^2 - 15$

Solution

There are three operations to do:, multiplication, squaring, and subtraction. The Order of Operations tells us to complete exponents before multiplication and then subtraction. Therefore, the solution is 85 and is found as follows:

$4 \times 5^2 - 15 = 4 \times 25 - 15 = 100 - 15 = 85$

Example 15.2.3

$36 \div 12 + (8 \div 2)$

Solution

There are three operations to do: divide, add, and divide within a parenthesis. The Order of Operations tells us to complete the division within parenthesis, then the division outside the parenthesis, and then the addition. Therefore, the solution is 7 and is found as follows:

$36 \div 12 + (8 \div 2) = 36 \div 12 + 4 = 3 + 4 = 7$

15.3: Appendix C- The Z-Distribution Table

The first column shows the positive version of each z-scores. The remaining columns report proportions in decimal form. The second column shows the proportion of scores from the z-score past the mean and to its furthest tail in the normal curve. The third column shows the proportion of scores from the z-score to its nearest tail of the normal curve. The fourth column shows the proportion of scores between the z-score and the mean (center) of the normal curve. To translate proportions to percents, multiply the proportion by 100.

Recommendation: It can be easiest to identify which column(s) you need by using the sketch, shade, and solve method outlined in Chapter 5.

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
0.00	0.5000	0.5000	0.0000
0.01	0.5040	0.4960	0.0040
0.02	0.5080	0.4920	0.0080
0.03	0.5120	0.4880	0.0120
0.04	0.5160	0.4840	0.0160
0.05	0.5199	0.4801	0.0199
0.06	0.5239	0.4761	0.0239
0.07	0.5279	0.4721	0.0279
0.08	0.5319	0.4681	0.0319
0.09	0.5359	0.4641	0.0359
0.10	0.5398	0.4602	0.0398
0.11	0.5438	0.4562	0.0438
0.12	0.5478	0.4522	0.0478
0.13	0.5517	0.4483	0.0517
0.14	0.5557	0.4443	0.0557
0.15	0.5596	0.4404	0.0596
0.16	0.5636	0.4364	0.0636
0.17	0.5675	0.4325	0.0675
0.18	0.5714	0.4286	0.0714
0.19	0.5753	0.4247	0.0753
0.20	0.5793	0.4207	0.0793
0.21	0.5832	0.4168	0.0832
0.22	0.5871	0.4129	0.0871
0.23	0.5910	0.4090	0.0910
0.24	0.5948	0.4052	0.0948
0.25	0.5987	0.4013	0.0987
0.26	0.6026	0.3974	0.1026

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
0.27	0.6064	0.3936	0.1064
0.28	0.6103	0.3897	0.1103
0.29	0.6141	0.3859	0.1141
0.30	0.6179	0.3821	0.1179
0.31	0.6217	0.3783	0.1217
0.32	0.6255	0.3745	0.1255
0.33	0.6293	0.3707	0.1293
0.34	0.6331	0.3669	0.1331
0.35	0.6368	0.3632	0.1368
0.36	0.6406	0.3594	0.1406
0.37	0.6443	0.3557	0.1443
0.38	0.6480	0.3520	0.1480
0.39	0.6517	0.3483	0.1517
0.40	0.6554	0.3446	0.1554
0.41	0.6591	0.3409	0.1591
0.42	0.6628	0.3372	0.1628
0.43	0.6664	0.3336	0.1664
0.44	0.6700	0.3300	0.1700
0.45	0.6736	0.3264	0.1736
0.46	0.6772	0.3228	0.1772
0.47	0.6808	0.3192	0.1808
0.48	0.6844	0.3156	0.1844
0.49	0.6879	0.3121	0.1879
0.50	0.6915	0.3085	0.1915
0.51	0.6950	0.3050	0.1950
0.52	0.6985	0.3015	0.1985
0.53	0.7019	0.2981	0.2019
0.54	0.7054	0.2946	0.2054
0.55	0.7088	0.2912	0.2088
0.56	0.7123	0.2877	0.2123
0.57	0.7157	0.2843	0.2157
0.58	0.7190	0.2810	0.2190
0.59	0.7224	0.2776	0.2224
0.60	0.7257	0.2743	0.2257

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
0.61	0.7291	0.2709	0.2291
0.62	0.7324	0.2676	0.2324
0.63	0.7357	0.2643	0.2357
0.64	0.7389	0.2611	0.2389
0.65	0.7422	0.2578	0.2422
0.66	0.7454	0.2546	0.2454
0.67	0.7486	0.2514	0.2486
0.68	0.7517	0.2483	0.2517
0.69	0.7549	0.2451	0.2549
0.70	0.7580	0.2420	0.2580
0.71	0.7611	0.2389	0.2611
0.72	0.7642	0.2358	0.2642
0.73	0.7673	0.2327	0.2673
0.74	0.7704	0.2296	0.2704
0.75	0.7734	0.2266	0.2734
0.76	0.7764	0.2236	0.2764
0.77	0.7794	0.2206	0.2794
0.78	0.7823	0.2177	0.2823
0.79	0.7852	0.2148	0.2852
0.80	0.7881	0.2119	0.2881
0.81	0.7910	0.2090	0.2910
0.82	0.7939	0.2061	0.2939
0.83	0.7967	0.2033	0.2967
0.84	0.7995	0.2005	0.2995
0.85	0.8023	0.1977	0.3023
0.86	0.8051	0.1949	0.3051
0.87	0.8078	0.1922	0.3078
0.88	0.8106	0.1894	0.3106
0.89	0.8133	0.1867	0.3133
0.90	0.8159	0.1841	0.3159
0.91	0.8186	0.1814	0.3186
0.92	0.8212	0.1788	0.3212
0.93	0.8238	0.1762	0.3238
0.94	0.8264	0.1736	0.3264

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
0.95	0.8289	0.1711	0.3289
0.96	0.8315	0.1685	0.3315
0.97	0.8340	0.1660	0.3340
0.98	0.8365	0.1635	0.3365
0.99	0.8389	0.1611	0.3389
1.00	0.8413	0.1587	0.3413
1.01	0.8438	0.1562	0.3438
1.02	0.8461	0.1539	0.3461
1.03	0.8485	0.1515	0.3485
1.04	0.8508	0.1492	0.3508
1.05	0.8531	0.1469	0.3531
1.06	0.8554	0.1446	0.3554
1.07	0.8577	0.1423	0.3577
1.08	0.8599	0.1401	0.3599
1.09	0.8621	0.1379	0.3621
1.10	0.8643	0.1357	0.3643
1.11	0.8665	0.1335	0.3665
1.12	0.8686	0.1314	0.3686
1.13	0.8708	0.1292	0.3708
1.14	0.8729	0.1271	0.3729
1.15	0.8749	0.1251	0.3749
1.16	0.8770	0.1230	0.3770
1.17	0.8790	0.1210	0.3790
1.18	0.8810	0.1190	0.3810
1.19	0.8830	0.1170	0.3830
1.20	0.8849	0.1151	0.3849
1.21	0.8869	0.1131	0.3869
1.22	0.8888	0.1112	0.3888
1.23	0.8907	0.1093	0.3907
1.24	0.8925	0.1075	0.3925
1.25	0.8944	0.1056	0.3944
1.26	0.8962	0.1038	0.3962
1.27	0.8980	0.1020	0.3980
1.28	0.8997	0.1003	0.3997

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
1.29	0.9015	0.0985	0.4015
1.30	0.9032	0.0968	0.4032
1.31	0.9049	0.0951	0.4049
1.32	0.9066	0.0934	0.4066
1.33	0.9082	0.0918	0.4082
1.34	0.9099	0.0901	0.4099
1.35	0.9115	0.0885	0.4115
1.36	0.9131	0.0869	0.4131
1.37	0.9147	0.0853	0.4147
1.38	0.9162	0.0838	0.4162
1.39	0.9177	0.0823	0.4177
1.40	0.9192	0.0808	0.4192
1.41	0.9207	0.0793	0.4207
1.42	0.9222	0.0778	0.4222
1.43	0.9236	0.0764	0.4236
1.44	0.9251	0.0749	0.4251
1.45	0.9265	0.0735	0.4265
1.46	0.9279	0.0721	0.4279
1.47	0.9292	0.0708	0.4292
1.48	0.9306	0.0694	0.4306
1.49	0.9319	0.0681	0.4319
1.50	0.9332	0.0668	0.4332
1.51	0.9345	0.0655	0.4345
1.52	0.9357	0.0643	0.4357
1.53	0.9370	0.0630	0.4370
1.54	0.9382	0.0618	0.4382
1.55	0.9394	0.0606	0.4394
1.56	0.9406	0.0594	0.4406
1.57	0.9418	0.0582	0.4418
1.58	0.9429	0.0571	0.4429
1.59	0.9441	0.0559	0.4441
1.60	0.9452	0.0548	0.4452
1.61	0.9463	0.0537	0.4463
1.62	0.9474	0.0526	0.4474

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
1.63	0.9484	0.0516	0.4484
1.64	0.9495	0.0505	0.4495
1.65	0.9505	0.0495	0.4505
1.66	0.9515	0.0485	0.4515
1.67	0.9525	0.0475	0.4525
1.68	0.9535	0.0465	0.4535
1.69	0.9545	0.0455	0.4545
1.70	0.9554	0.0446	0.4554
1.71	0.9564	0.0436	0.4564
1.72	0.9573	0.0427	0.4573
1.73	0.9582	0.0418	0.4582
1.74	0.9591	0.0409	0.4591
1.75	0.9599	0.0401	0.4599
1.76	0.9608	0.0392	0.4608
1.77	0.9616	0.0384	0.4616
1.78	0.9625	0.0375	0.4625
1.79	0.9633	0.0367	0.4633
1.80	0.9641	0.0359	0.4641
1.81	0.9649	0.0351	0.4649
1.82	0.9656	0.0344	0.4656
1.83	0.9664	0.0336	0.4664
1.84	0.9671	0.0329	0.4671
1.85	0.9678	0.0322	0.4678
1.86	0.9686	0.0314	0.4686
1.87	0.9693	0.0307	0.4693
1.88	0.9699	0.0301	0.4699
1.89	0.9706	0.0294	0.4706
1.90	0.9713	0.0287	0.4713
1.91	0.9719	0.0281	0.4719
1.92	0.9726	0.0274	0.4726
1.93	0.9732	0.0268	0.4732
1.94	0.9738	0.0262	0.4738
1.95	0.9744	0.0256	0.4744
1.96	0.9750	0.0250	0.4750

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
1.97	0.9756	0.0244	0.4756
1.98	0.9761	0.0239	0.4761
1.99	0.9767	0.0233	0.4767
2.00	0.9772	0.0228	0.4772
2.01	0.9778	0.0222	0.4778
2.02	0.9783	0.0217	0.4783
2.03	0.9788	0.0212	0.4788
2.04	0.9793	0.0207	0.4793
2.05	0.9798	0.0202	0.4798
2.06	0.9803	0.0197	0.4803
2.07	0.9808	0.0192	0.4808
2.08	0.9812	0.0188	0.4812
2.09	0.9817	0.0183	0.4817
2.10	0.9821	0.0179	0.4821
2.11	0.9826	0.0174	0.4826
2.12	0.9830	0.0170	0.4830
2.13	0.9834	0.0166	0.4834
2.14	0.9838	0.0162	0.4838
2.15	0.9842	0.0158	0.4842
2.16	0.9846	0.0154	0.4846
2.17	0.9850	0.0150	0.4850
2.18	0.9854	0.0146	0.4854
2.19	0.9857	0.0143	0.4857
2.20	0.9861	0.0139	0.4861
2.21	0.9864	0.0136	0.4864
2.22	0.9868	0.0132	0.4868
2.23	0.9871	0.0129	0.4871
2.24	0.9875	0.0125	0.4875
2.25	0.9878	0.0122	0.4878
2.26	0.9881	0.0119	0.4881
2.27	0.9884	0.0116	0.4884
2.28	0.9887	0.0113	0.4887
2.29	0.9890	0.0110	0.4890
2.30	0.9893	0.0107	0.4893

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
2.31	0.9896	0.0104	0.4896
2.32	0.9898	0.0102	0.4898
2.33	0.9901	0.0099	0.4901
2.34	0.9904	0.0096	0.4904
2.35	0.9906	0.0094	0.4906
2.36	0.9909	0.0091	0.4909
2.37	0.9911	0.0089	0.4911
2.38	0.9913	0.0087	0.4913
2.39	0.9916	0.0084	0.4916
2.40	0.9918	0.0082	0.4918
2.41	0.9920	0.0080	0.4920
2.42	0.9922	0.0078	0.4922
2.43	0.9925	0.0075	0.4925
2.44	0.9927	0.0073	0.4927
2.45	0.9929	0.0071	0.4929
2.46	0.9931	0.0069	0.4931
2.47	0.9932	0.0068	0.4932
2.48	0.9934	0.0066	0.4934
2.49	0.9936	0.0064	0.4936
2.50	0.9938	0.0062	0.4938
2.51	0.9940	0.0060	0.4940
2.52	0.9941	0.0059	0.4941
2.53	0.9943	0.0057	0.4943
2.54	0.9945	0.0055	0.4945
2.55	0.9946	0.0054	0.4946
2.56	0.9948	0.0052	0.4948
2.57	0.9949	0.0051	0.4949
2.58	0.9951	0.0049	0.4951
2.59	0.9952	0.0048	0.4952
2.60	0.9953	0.0047	0.4953
2.61	0.9955	0.0045	0.4955
2.62	0.9956	0.0044	0.4956
2.63	0.9957	0.0043	0.4957
2.64	0.9959	0.0041	0.4959

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
2.65	0.9960	0.0040	0.4960
2.66	0.9961	0.0039	0.4961
2.67	0.9962	0.0038	0.4962
2.68	0.9963	0.0037	0.4963
2.69	0.9964	0.0036	0.4964
2.70	0.9965	0.0035	0.4965
2.71	0.9966	0.0034	0.4966
2.72	0.9967	0.0033	0.4967
2.73	0.9968	0.0032	0.4968
2.74	0.9969	0.0031	0.4969
2.75	0.9970	0.0030	0.4970
2.76	0.9971	0.0029	0.4971
2.77	0.9972	0.0028	0.4972
2.78	0.9973	0.0027	0.4973
2.79	0.9974	0.0026	0.4974
2.80	0.9974	0.0026	0.4974
2.81	0.9975	0.0025	0.4975
2.82	0.9976	0.0024	0.4976
2.83	0.9977	0.0023	0.4977
2.84	0.9977	0.0023	0.4977
2.85	0.9978	0.0022	0.4978
2.86	0.9979	0.0021	0.4979
2.87	0.9979	0.0021	0.4979
2.88	0.9980	0.0020	0.4980
2.89	0.9981	0.0019	0.4981
2.90	0.9981	0.0019	0.4981
2.91	0.9982	0.0018	0.4982
2.92	0.9982	0.0018	0.4982
2.93	0.9983	0.0017	0.4983
2.94	0.9984	0.0016	0.4984
2.95	0.9984	0.0016	0.4984
2.96	0.9985	0.0015	0.4985
2.97	0.9985	0.0015	0.4985
2.98	0.9986	0.0014	0.4986

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
2.99	0.9986	0.0014	0.4986
3.00	0.9987	0.0013	0.4987
3.01	0.9987	0.0013	0.4987
3.02	0.9987	0.0013	0.4987
3.03	0.9988	0.0012	0.4988
3.04	0.9988	0.0012	0.4988
3.05	0.9989	0.0011	0.4989
3.06	0.9989	0.0011	0.4989
3.07	0.9989	0.0011	0.4989
3.08	0.9990	0.0010	0.4990
3.09	0.9990	0.0010	0.4990
3.10	0.9990	0.0010	0.4990
3.11	0.9991	0.0009	0.4991
3.12	0.9991	0.0009	0.4991
3.13	0.9991	0.0009	0.4991
3.14	0.9992	0.0008	0.4992
3.15	0.9992	0.0008	0.4992
3.16	0.9992	0.0008	0.4992
3.17	0.9992	0.0008	0.4992
3.18	0.9993	0.0007	0.4993
3.19	0.9993	0.0007	0.4993
3.20	0.9993	0.0007	0.4993
3.21	0.9993	0.0007	0.4993
3.22	0.9994	0.0006	0.4994
3.23	0.9994	0.0006	0.4994
3.24	0.9994	0.0006	0.4994
3.25	0.9994	0.0006	0.4994
3.26	0.9994	0.0006	0.4994
3.27	0.9995	0.0005	0.4995
3.28	0.9995	0.0005	0.4995
3.29	0.9995	0.0005	0.4995
3.30	0.9995	0.0005	0.4995
3.31	0.9995	0.0005	0.4995
3.32	0.9995	0.0005	0.4995

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
3.33	0.9996	0.0004	0.4996
3.34	0.9996	0.0004	0.4996
3.35	0.9996	0.0004	0.4996
3.36	0.9996	0.0004	0.4996
3.37	0.9996	0.0004	0.4996
3.38	0.9996	0.0004	0.4996
3.39	0.9997	0.0003	0.4997
3.40	0.9997	0.0003	0.4997
3.41	0.9997	0.0003	0.4997
3.42	0.9997	0.0003	0.4997
3.43	0.9997	0.0003	0.4997
3.44	0.9997	0.0003	0.4997
3.45	0.9997	0.0003	0.4997
3.46	0.9997	0.0003	0.4997
3.47	0.9997	0.0003	0.4997
3.48	0.9997	0.0003	0.4997
3.49	0.9998	0.0002	0.4998
3.50	0.9998	0.0002	0.4998
3.51	0.9998	0.0002	0.4998
3.52	0.9998	0.0002	0.4998
3.53	0.9998	0.0002	0.4998
3.54	0.9998	0.0002	0.4998
3.55	0.9998	0.0002	0.4998
3.56	0.9998	0.0002	0.4998
3.57	0.9998	0.0002	0.4998
3.58	0.9998	0.0002	0.4998
3.59	0.9998	0.0002	0.4998
3.60	0.9998	0.0002	0.4998
3.61	0.9998	0.0002	0.4998
3.62	0.9999	0.0001	0.4999
3.63	0.9999	0.0001	0.4999
3.64	0.9999	0.0001	0.4999
3.65	0.9999	0.0001	0.4999
3.66	0.9999	0.0001	0.4999

Z-Score	Proportion in the Body	Proportion in the Tail	Proportion between the Mean and the Z
3.67	0.9999	0.0001	0.4999
3.68	0.9999	0.0001	0.4999
3.69	0.9999	0.0001	0.4999
3.70	0.9999	0.0001	0.4999
3.71	0.9999	0.0001	0.4999
3.72	0.9999	0.0001	0.4999
3.73	0.9999	0.0001	0.4999
3.74	0.9999	0.0001	0.4999
3.75	0.9999	0.0001	0.4999
3.76	0.9999	0.0001	0.4999
3.77	0.9999	0.0001	0.4999
3.78	0.9999	0.0001	0.4999
3.79	0.9999	0.0001	0.4999
3.80	0.9999	0.0001	0.4999
3.81	0.9999	0.0001	0.4999
3.82	0.9999	0.0001	0.4999
3.83	0.9999	0.0001	0.4999
3.84	0.9999	0.0001	0.4999
3.85	0.9999	0.0001	0.4999
3.86	0.9999	0.0001	0.4999
3.87	0.9999	0.0001	0.4999
3.88	0.9999	0.0001	0.4999
3.89	0.9999	0.0001	0.4999
3.90	1.0000	0.0000	0.5000
3.91	1.0000	0.0000	0.5000
3.92	1.0000	0.0000	0.5000
3.93	1.0000	0.0000	0.5000
3.94	1.0000	0.0000	0.5000
3.95	1.0000	0.0000	0.5000
3.96	1.0000	0.0000	0.5000
3.97	1.0000	0.0000	0.5000
3.98	1.0000	0.0000	0.5000
3.99	1.0000	0.0000	0.5000
4.00	1.0000	0.0000	0.5000

15.4: Appendix D- The t-Tables

The critical values shown are for use with non-directional (two-tailed) and directional (one-tailed) *t*-tests.

		two-tailed test		one-tailed test	
Degrees of Freedom	alpha level:	α = 0.05	α = 0.01	α = 0.05	α = 0.01
	1	12.706	63.657	6.314	31.821
	2	4.303	9.925	2.920	6.965
	3	3.182	5.841	2.353	4.541
	4	2.776	4.604	2.132	3.747
	5	2.571	4.032	2.015	3.365
	6	2.447	3.707	1.943	3.143
	7	2.365	3.499	1.895	2.998
	8	2.306	3.355	1.860	2.896
	9	2.262	3.250	1.833	2.821
	10	2.228	3.169	1.812	2.764
	11	2.201	3.106	1.796	2.718
	12	2.179	3.055	1.782	2.681
	13	2.160	3.012	1.771	2.650
	14	2.145	2.977	1.761	2.624
	15	2.131	2.947	1.753	2.602
	16	2.120	2.921	1.746	2.583
	17	2.110	2.898	1.740	2.567
	18	2.101	2.878	1.734	2.552
	19	2.093	2.861	1.729	2.539
	20	2.086	2.845	1.725	2.528
	21	2.080	2.831	1.721	2.518
	22	2.074	2.819	1.717	2.508
	23	2.069	2.807	1.714	2.500
	24	2.064	2.797	1.711	2.492
	25	2.060	2.787	1.708	2.485
	26	2.056	2.779	1.706	2.479
	27	2.052	2.771	1.703	2.473
	28	2.048	2.763	1.701	2.467
	29	2.045	2.756	1.699	2.462
	30	2.042	2.750	1.697	2.457
	31	2.040	2.744	1.696	2.453

	two-tailed test		one-tailed test	
32	2.037	2.738	1.694	2.449
33	2.035	2.733	1.692	2.445
34	2.032	2.728	1.691	2.441
35	2.030	2.724	1.690	2.438
36	2.028	2.719	1.688	2.434
37	2.026	2.715	1.687	2.431
38	2.024	2.712	1.686	2.429
39	2.023	2.708	1.685	2.426
40	2.021	2.704	1.684	2.423

15.5: Appendix E- The F-Tables

The tables show the critical values for a simple independent-groups ANOVA.

When using repeated-measures ANOVA, use df_e in place of df_w.

When using ANOVA as part of a regression, use df_M in place of df_b and df_e in place of df_w.

When using α = 0.05, use standard values.

When using α = 0.01, use bolded values.

	Degrees of Freedom Between (df_b)								
Degrees of Freedom Within (df_w)		1	2	3	4	5	6	7	8
	1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883
		4052.181	**4999.500**	**5403.352**	**5624.583**	**5763.650**	**5858.986**	**5928.356**	**5981.070**
	2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371
		98.503	**99.000**	**99.166**	**99.249**	**99.299**	**99.333**	**99.356**	**99.374**
	3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845
		34.116	**30.817**	**29.457**	**28.710**	**28.237**	**27.911**	**27.672**	**27.489**
	4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041
		21.198	**18.000**	**16.694**	**15.977**	**15.522**	**15.207**	**14.976**	**14.799**
	5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818
		16.258	**13.274**	**12.060**	**11.392**	**10.967**	**10.672**	**10.456**	**10.289**
	6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147
		13.745	**10.925**	**9.780**	**9.148**	**8.746**	**8.466**	**8.260**	**8.102**
	7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726
		12.246	**9.547**	**8.451**	**7.847**	**7.460**	**7.191**	**6.993**	**6.840**
	8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438
		11.259	**8.649**	**7.591**	**7.006**	**6.632**	**6.371**	**6.178**	**6.029**
	9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230
		10.561	**8.022**	**6.992**	**6.422**	**6.057**	**5.802**	**5.613**	**5.467**
	10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072
		10.044	**7.559**	**6.552**	**5.994**	**5.636**	**5.386**	**5.200**	**5.057**
	11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948
		9.646	**7.206**	**6.217**	**5.668**	**5.316**	**5.069**	**4.886**	**4.744**
	12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849
		9.330	**6.927**	**5.953**	**5.412**	**5.064**	**4.821**	**4.640**	**4.499**
	13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767
		9.074	**6.701**	**5.739**	**5.205**	**4.862**	**4.620**	**4.441**	**4.302**
	14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699

Degrees of Freedom Between (df_b)								
	8.862	**6.515**	**5.564**	**5.035**	**4.695**	**4.456**	**4.278**	**4.140**
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641
	8.683	**6.359**	**5.417**	**4.893**	**4.556**	**4.318**	**4.142**	**4.004**
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591
	8.531	**6.226**	**5.292**	**4.773**	**4.437**	**4.202**	**4.026**	**3.890**
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548
	8.400	**6.112**	**5.185**	**4.669**	**4.336**	**4.102**	**3.927**	**3.791**
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510
	8.285	**6.013**	**5.092**	**4.579**	**4.248**	**4.015**	**3.841**	**3.705**
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477
	8.185	**5.926**	**5.010**	**4.500**	**4.171**	**3.939**	**3.765**	**3.631**
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447
	8.096	**5.849**	**4.938**	**4.431**	**4.103**	**3.871**	**3.699**	**3.564**
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420
	8.017	**5.780**	**4.874**	**4.369**	**4.042**	**3.812**	**3.640**	**3.506**
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397
	7.945	**5.719**	**4.817**	**4.313**	**3.988**	**3.758**	**3.587**	**3.453**
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375
	7.881	**5.664**	**4.765**	**4.264**	**3.939**	**3.710**	**3.539**	**3.406**
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355
	7.823	**5.614**	**4.718**	**4.218**	**3.895**	**3.667**	**3.496**	**3.363**
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337
	7.770	**5.568**	**4.675**	**4.177**	**3.855**	**3.627**	**3.457**	**3.324**
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321
	7.721	**5.526**	**4.637**	**4.140**	**3.818**	**3.591**	**3.421**	**3.288**
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305
	7.677	**5.488**	**4.601**	**4.106**	**3.785**	**3.558**	**3.388**	**3.256**
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291
	7.636	**5.453**	**4.568**	**4.074**	**3.754**	**3.528**	**3.358**	**3.226**
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278
	7.598	**5.420**	**4.538**	**4.045**	**3.725**	**3.499**	**3.330**	**3.198**
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266
	7.562	**5.390**	**4.510**	**4.018**	**3.699**	**3.473**	**3.304**	**3.173**
31	4.160	3.305	2.911	2.679	2.523	2.409	2.323	2.255

Degrees of Freedom Between (df_b)								
	7.530	**5.362**	**4.484**	**3.993**	**3.675**	**3.449**	**3.281**	**3.149**
32	4.149	3.295	2.901	2.668	2.512	2.399	2.313	2.244
	7.499	**5.336**	**4.459**	**3.969**	**3.652**	**3.427**	**3.258**	**3.127**
33	4.139	3.285	2.892	2.659	2.503	2.389	2.303	2.235
	7.471	**5.312**	**4.437**	**3.948**	**3.630**	**3.406**	**3.238**	**3.106**
34	4.130	3.276	2.883	2.650	2.494	2.380	2.294	2.225
	7.444	**5.289**	**4.416**	**3.927**	**3.611**	**3.386**	**3.218**	**3.087**
35	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217
	7.419	**5.268**	**4.396**	**3.908**	**3.592**	**3.368**	**3.200**	**3.069**
36	4.113	3.259	2.866	2.634	2.477	2.364	2.277	2.209
	7.396	**5.248**	**4.377**	**3.890**	**3.574**	**3.351**	**3.183**	**3.052**
37	4.105	3.252	2.859	2.626	2.470	2.356	2.270	2.201
	7.373	**5.229**	**4.360**	**3.873**	**3.558**	**3.334**	**3.167**	**3.036**
38	4.098	3.245	2.852	2.619	2.463	2.349	2.262	2.194
	7.353	**5.211**	**4.343**	**3.858**	**3.542**	**3.319**	**3.152**	**3.021**
39	4.091	3.238	2.845	2.612	2.456	2.342	2.255	2.187
	7.333	**5.194**	**4.327**	**3.843**	**3.528**	**3.305**	**3.137**	**3.006**
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180
	7.314	**5.179**	**4.313**	**3.828**	**3.514**	**3.291**	**3.124**	**2.993**
41	4.079	3.226	2.833	2.600	2.443	2.330	2.243	2.174
	7.296	**5.163**	**4.299**	**3.815**	**3.501**	**3.278**	**3.111**	**2.980**
42	4.073	3.220	2.827	2.594	2.438	2.324	2.237	2.168
	7.280	**5.149**	**4.285**	**3.802**	**3.488**	**3.266**	**3.099**	**2.968**
43	4.067	3.214	2.822	2.589	2.432	2.318	2.232	2.163
	7.264	**5.136**	**4.273**	**3.790**	**3.476**	**3.254**	**3.087**	**2.957**
44	4.062	3.209	2.816	2.584	2.427	2.313	2.226	2.157
	7.248	**5.123**	**4.261**	**3.778**	**3.465**	**3.243**	**3.076**	**2.946**
45	4.057	3.204	2.812	2.579	2.422	2.308	2.221	2.152
	7.234	**5.110**	**4.249**	**3.767**	**3.454**	**3.232**	**3.066**	**2.935**
46	4.052	3.200	2.807	2.574	2.417	2.304	2.216	2.147
	7.220	**5.099**	**4.238**	**3.757**	**3.444**	**3.222**	**3.056**	**2.925**
47	4.047	3.195	2.802	2.570	2.413	2.299	2.212	2.143
	7.207	**5.087**	**4.228**	**3.747**	**3.434**	**3.213**	**3.046**	**2.916**
48	4.043	3.191	2.798	2.565	2.409	2.295	2.207	2.138

Degrees of Freedom Between (df_b)								
	7.194	**5.077**	**4.218**	**3.737**	**3.425**	**3.204**	**3.037**	**2.907**
49	4.038	3.187	2.794	2.561	2.404	2.290	2.203	2.134
	7.182	**5.066**	**4.208**	**3.728**	**3.416**	**3.195**	**3.028**	**2.898**
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130
	7.171	**5.057**	**4.199**	**3.720**	**3.408**	**3.186**	**3.020**	**2.890**
51	4.030	3.179	2.786	2.553	2.397	2.283	2.195	2.126
	7.159	**5.047**	**4.191**	**3.711**	**3.400**	**3.178**	**3.012**	**2.882**
52	4.027	3.175	2.783	2.550	2.393	2.279	2.192	2.122
	7.149	**5.038**	**4.182**	**3.703**	**3.392**	**3.171**	**3.005**	**2.874**
53	4.023	3.172	2.779	2.546	2.389	2.275	2.188	2.119
	7.139	**5.030**	**4.174**	**3.695**	**3.384**	**3.163**	**2.997**	**2.867**
54	4.020	3.168	2.776	2.543	2.386	2.272	2.185	2.115
	7.129	**5.021**	**4.167**	**3.688**	**3.377**	**3.156**	**2.990**	**2.860**
55	4.016	3.165	2.773	2.540	2.383	2.269	2.181	2.112
	7.119	**5.013**	**4.159**	**3.681**	**3.370**	**3.149**	**2.983**	**2.853**
56	4.013	3.162	2.769	2.537	2.380	2.266	2.178	2.109
	7.110	**5.006**	**4.152**	**3.674**	**3.363**	**3.143**	**2.977**	**2.847**
57	4.010	3.159	2.766	2.534	2.377	2.263	2.175	2.106
	7.102	**4.998**	**4.145**	**3.667**	**3.357**	**3.136**	**2.971**	**2.841**
58	4.007	3.156	2.764	2.531	2.374	2.260	2.172	2.103
	7.093	**4.991**	**4.138**	**3.661**	**3.351**	**3.130**	**2.965**	**2.835**
59	4.004	3.153	2.761	2.528	2.371	2.257	2.169	2.100
	7.085	**4.984**	**4.132**	**3.655**	**3.345**	**3.124**	**2.959**	**2.829**
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097
	7.077	**4.977**	**4.126**	**3.649**	**3.339**	**3.119**	**2.953**	**2.823**
61	3.998	3.148	2.755	2.523	2.366	2.251	2.164	2.094
	7.070	**4.971**	**4.120**	**3.643**	**3.333**	**3.113**	**2.948**	**2.818**
62	3.996	3.145	2.753	2.520	2.363	2.249	2.161	2.092
	7.062	**4.965**	**4.114**	**3.638**	**3.328**	**3.108**	**2.942**	**2.813**
63	3.993	3.143	2.751	2.518	2.361	2.246	2.159	2.089
	7.055	**4.959**	**4.109**	**3.632**	**3.323**	**3.103**	**2.937**	**2.808**
64	3.991	3.140	2.748	2.515	2.358	2.244	2.156	2.087
	7.048	**4.953**	**4.103**	**3.627**	**3.318**	**3.098**	**2.932**	**2.803**
65	3.989	3.138	2.746	2.513	2.356	2.242	2.154	2.084

	Degrees of Freedom Between (df_b)							
	7.042	**4.947**	**4.098**	**3.622**	**3.313**	**3.093**	**2.928**	**2.798**
66	3.986	3.136	2.744	2.511	2.354	2.239	2.152	2.082
	7.035	**4.942**	**4.093**	**3.618**	**3.308**	**3.088**	**2.923**	**2.793**
67	3.984	3.134	2.742	2.509	2.352	2.237	2.150	2.080
	7.029	**4.937**	**4.088**	**3.613**	**3.304**	**3.084**	**2.919**	**2.789**
68	3.982	3.132	2.740	2.507	2.350	2.235	2.148	2.078
	7.023	**4.932**	**4.083**	**3.608**	**3.299**	**3.080**	**2.914**	**2.785**
69	3.980	3.130	2.737	2.505	2.348	2.233	2.145	2.076
	7.017	**4.927**	**4.079**	**3.604**	**3.295**	**3.075**	**2.910**	**2.781**
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074
	7.011	**4.922**	**4.074**	**3.600**	**3.291**	**3.071**	**2.906**	**2.777**
71	3.976	3.126	2.734	2.501	2.344	2.229	2.142	2.072
	7.006	**4.917**	**4.070**	**3.596**	**3.287**	**3.067**	**2.902**	**2.773**
72	3.974	3.124	2.732	2.499	2.342	2.227	2.140	2.070
	7.001	**4.913**	**4.066**	**3.591**	**3.283**	**3.063**	**2.898**	**2.769**
73	3.972	3.122	2.730	2.497	2.340	2.226	2.138	2.068
	6.995	**4.908**	**4.062**	**3.588**	**3.279**	**3.060**	**2.895**	**2.765**
74	3.970	3.120	2.728	2.495	2.338	2.224	2.136	2.066
	6.990	**4.904**	**4.058**	**3.584**	**3.275**	**3.056**	**2.891**	**2.762**
75	3.968	3.119	2.727	2.494	2.337	2.222	2.134	2.064
	6.985	**4.900**	**4.054**	**3.580**	**3.272**	**3.052**	**2.887**	**2.758**
76	3.967	3.117	2.725	2.492	2.335	2.220	2.133	2.063
	6.981	**4.896**	**4.050**	**3.577**	**3.268**	**3.049**	**2.884**	**2.755**
77	3.965	3.115	2.723	2.490	2.333	2.219	2.131	2.061
	6.976	**4.892**	**4.047**	**3.573**	**3.265**	**3.046**	**2.881**	**2.751**
78	3.963	3.114	2.722	2.489	2.332	2.217	2.129	2.059
	6.971	**4.888**	**4.043**	**3.570**	**3.261**	**3.042**	**2.877**	**2.748**
79	3.962	3.112	2.720	2.487	2.330	2.216	2.128	2.058
	6.967	**4.884**	**4.040**	**3.566**	**3.258**	**3.039**	**2.874**	**2.745**
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056
	6.963	**4.881**	**4.036**	**3.563**	**3.255**	**3.036**	**2.871**	**2.742**

15.6: Appendix F- The q-Tables

The tables show the Studentized Range q-values needed when hand-calculating HSD post-hoc comparisons.

	Number of Groups (k)								
Degrees of Freedom Within (df_w)		**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	1	17.969	26.976	32.819	37.082	40.408	43.119	45.397	47.357
	2	6.085	8.331	9.798	10.881	11.734	12.435	13.027	13.538
	3	4.501	5.910	6.825	7.502	8.037	8.478	8.852	9.177
	4	3.927	5.040	5.757	6.287	6.707	7.053	7.347	7.602
	5	3.635	4.602	5.219	5.673	6.033	6.330	6.582	6.801
	6	3.461	4.339	4.896	5.305	5.629	5.895	6.122	6.319
	7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.998
	8	3.261	4.041	4.529	4.886	5.167	5.399	5.596	5.767
	9	3.199	3.949	4.415	4.755	5.024	5.244	5.432	5.595
	10	3.151	3.877	4.327	4.654	4.912	5.124	5.304	5.461
	11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353
	12	3.081	3.773	4.199	4.508	4.748	4.947	5.116	5.263
	13	3.055	3.734	4.151	4.453	4.690	4.884	5.049	5.192
	14	3.033	3.701	4.111	4.407	4.639	4.829	4.990	5.130
	15	3.014	3.673	4.076	4.367	4.595	4.782	4.940	5.077
	16	2.998	3.649	4.046	4.333	4.557	4.741	4.896	5.031
	17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991
	18	2.971	3.609	3.997	4.276	4.494	4.673	4.824	4.955
	19	2.960	3.593	3.977	4.253	4.469	4.645	4.794	4.924
	20	2.950	3.578	3.958	4.232	4.445	4.620	4.768	4.895
	21	2.941	3.565	3.942	4.213	4.424	4.597	4.744	4.870
	22	2.933	3.553	3.927	4.196	4.406	4.577	4.722	4.847
	23	2.926	3.542	3.914	4.181	4.388	4.558	4.702	4.826
	24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807
	25	2.913	3.523	3.890	4.153	4.358	4.526	4.667	4.789
	26	2.907	3.514	3.880	4.142	4.345	4.512	4.652	4.773
	27	2.902	3.506	3.870	4.131	4.333	4.498	4.638	4.758
	28	2.897	3.499	3.861	4.120	4.322	4.486	4.625	4.745
	29	2.892	3.493	3.853	4.111	4.311	4.475	4.613	4.732
	30	2.888	3.487	3.845	4.102	4.302	4.464	4.601	4.720
	31	2.884	3.481	3.838	4.094	4.292	4.454	4.591	4.709

	32	2.881	3.475	3.832	4.086	4.284	4.445	4.581	4.698
	33	2.877	3.470	3.825	4.079	4.276	4.437	4.572	4.689
	34	2.874	3.465	3.820	4.072	4.268	4.428	4.563	4.680
	35	2.871	3.461	3.814	4.066	4.261	4.421	4.555	4.671
	36	2.868	3.457	3.809	4.060	4.255	4.414	4.547	4.663
	37	2.866	3.453	3.804	4.054	4.249	4.407	4.540	4.655
	38	2.863	3.449	3.799	4.049	4.243	4.400	4.533	4.648
	39	2.861	3.446	3.795	4.044	4.237	4.394	4.527	4.641
	40	2.858	3.442	3.791	4.039	4.232	4.389	4.521	4.635
	41	2.856	3.439	3.787	4.035	4.227	4.383	4.515	4.628
	42	2.854	3.436	3.783	4.030	4.222	4.378	4.509	4.622
	43	2.852	3.433	3.779	4.026	4.217	4.373	4.504	4.617
	44	2.850	3.430	3.776	4.022	4.213	4.368	4.499	4.611
	45	2.848	3.428	3.773	4.018	4.209	4.364	4.494	4.606
	46	2.847	3.425	3.770	4.015	4.205	4.359	4.489	4.601
	47	2.845	3.423	3.767	4.011	4.201	4.355	4.485	4.597
	48	2.844	3.420	3.764	4.008	4.197	4.351	4.481	4.592
	49	2.842	3.418	3.761	4.005	4.194	4.347	4.477	4.588
	50	2.841	3.416	3.758	4.002	4.190	4.344	4.473	4.584
	51	2.839	3.414	3.756	3.999	4.187	4.340	4.469	4.580

15.7: Appendix G- The r-Tables

The critical values shown are for use with non-directional (two-tailed) and directional (one-tailed) bivariate correlations (r). Degrees of Freedom $(df) = n - 2$

		two-tailed test		one-tailed test	
	alpha level:	α = 0.05	α = 0.01	α = 0.05	α = 0.01
Degrees of Freedom	1	.997	1.000	.988	1.000
	2	.950	.99	.900	.980
	3	.878	.959	.805	.934
	4	.811	.917	.729	.882
	5	.754	.875	.669	.833
	6	.707	.834	.621	.789
	7	.666	.798	.582	.760
	8	.632	.765	.549	.716
	9	.602	.735	.521	.685
	10	.576	.708	.497	.658
	11	553	.684	.476	.634
	12	.532	.661	.458	.612
	13	.514	.641	.441	.592
	14	.497	.623	.426	.574
	15	.482	.606	.412	.558
	16	.468	.590	.400	.542
	17	.456	.575	.389	.528
	18	.444	.561	.378	.516
	19	.433	.549	.369	.503
	20	.423	.537	.36	.492
	21	.413	.526	.352	.482
	22	.404	.515	.344	.472
	23	.396	.505	.337	.462
	24	.388	.496	.33	.453
	25	.381	.487	.323	.445
	26	.374	.479	.317	.437
	27	.367	.471	.311	.430
	28	.361	.463	.306	.423
	29	.355	.456	.301	.416
	30	.349	.449	.296	.409

Index

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Statistics: Open for Everyone (Peter)

Webpages: 137

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 100% (137 pages)

By Page

- Statistics: Open for Everyone (Peter) - *CC BY-NC-SA 4.0*
  - Front Matter - *CC BY-NC-SA 4.0*
    - TitlePage - *CC BY-NC-SA 4.0*
    - InfoPage - *CC BY-NC-SA 4.0*
    - Table of Contents - *CC BY-NC-SA 4.0*
    - Licensing - *CC BY-NC-SA 4.0*
  - 1: Introduction to Statistics - *CC BY-NC-SA 4.0*
    - 1.1: What is Statistics? - *CC BY-NC-SA 4.0*
    - 1.2: Tenets of Statistics - *CC BY-NC-SA 4.0*
    - 1.3: Types of Variables - *CC BY-NC-SA 4.0*
    - 1.4: Types of Statistics - *CC BY-NC-SA 4.0*
    - 1.5: Statistics and Science - *CC BY-NC-SA 4.0*
    - 1.6: How to Succeed in Statistics - *CC BY-NC-SA 4.0*
  - 2: Summarizing Data Visually - *CC BY-NC-SA 4.0*
    - 2.1: Visually Summarizing Data Univariately - *CC BY-NC-SA 4.0*
    - 2.2: How to Read a Data Set - *CC BY-NC-SA 4.0*
    - 2.3: The Value of Visual Summaries - *CC BY-NC-SA 4.0*
    - 2.4: Frequency Distribution Tables - *CC BY-NC-SA 4.0*
    - 2.5: Histograms - *CC BY-NC-SA 4.0*
    - 2.6: Frequency Polygons - *CC BY-NC-SA 4.0*
    - 2.7: Bar Graphs - *CC BY-NC-SA 4.0*
    - 2.8: Using SPSS - *CC BY-NC-SA 4.0*
    - 2.9: Chapter References - *CC BY-NC-SA 4.0*
  - 3: Measures of Central Tendency - *CC BY-NC-SA 4.0*
    - 3.1: The Goal of Summaries - *CC BY-NC-SA 4.0*
    - 3.2: Mode - *CC BY-NC-SA 4.0*
    - 3.3: Median - *CC BY-NC-SA 4.0*
    - 3.4: Mean - *CC BY-NC-SA 4.0*
    - 3.5: Reporting Results - *CC BY-NC-SA 4.0*
    - 3.6: Summary - *CC BY-NC-SA 4.0*
    - 3.7: Chapter References - *CC BY-NC-SA 4.0*
  - 4: Measures of Variability - *CC BY-NC-SA 4.0*
    - 4.1: Variability - *CC BY-NC-SA 4.0*
    - 4.2: Range - *CC BY-NC-SA 4.0*
    - 4.3: Standard Deviation - *CC BY-NC-SA 4.0*
    - 4.4: Summarizing Results in Sentences - *CC BY-NC-SA 4.0*
    - 4.5: Variance - *CC BY-NC-SA 4.0*
    - 4.6: Choosing a Measure of Variability - *CC BY-NC-SA 4.0*
    - 4.7: Using SPSS - *CC BY-NC-SA 4.0*
  - 5: The Normal Curve, z-Scores, and Probability - *CC BY-NC-SA 4.0*
    - 5.1: The Normal Curve, z-Scores, and Probability - *CC BY-NC-SA 4.0*
    - 5.2: The Normal Distribution Curve - *CC BY-NC-SA 4.0*
    - 5.3: z-Scores - *CC BY-NC-SA 4.0*
    - 5.4: Finding Probabilities - *CC BY-NC-SA 4.0*
    - 5.5: Probability and Inferential Statistics - *CC BY-NC-SA 4.0*
  - 6: The Foundations of Hypothesis Testing - *CC BY-NC-SA 4.0*
    - 6.1: Developing Hypotheses - *CC BY-NC-SA 4.0*
    - 6.2: One-Tailed vs. Two-Tailed Tests - *CC BY-NC-SA 4.0*
    - 6.3: Sampling Distributions - *CC BY-NC-SA 4.0*
    - 6.4: Hypothesis Testing - *CC BY-NC-SA 4.0*
    - 6.5: Errors and Statistical Significance - *CC BY-NC-SA 4.0*
    - 6.6: Choosing Statistical Tests - *CC BY-NC-SA 4.0*
    - 6.7: References - *CC BY-NC-SA 4.0*
  - 7: One Sample t-Tests - *CC BY-NC-SA 4.0*
    - 7.1: t-Statistics and the One Sample t-tests - *CC BY-NC-SA 4.0*
    - 7.2: The Logic of a t-Distribution - *CC BY-NC-SA 4.0*
    - 7.3: The One Sample t-Test Formula - *CC BY-NC-SA 4.0*
    - 7.4: Reporting Results - *CC BY-NC-SA 4.0*
    - 7.5: Practice Example of how to Compute t when Descriptive Statistics are Provided - *CC BY-NC-SA 4.0*

Detailed Licensing

Overview

Title: Statistics: Open for Everyone (Peter)

Webpages: 137

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 100% (137 pages)

By Page