2

CENTRAL TENDENCY AND VARIABILITY

LEARNING OBJECTIVES

- Create and interpret frequency distribution graphs and tables.
- Identify when to use the mean, median, or mode when describing a distribution's central tendency.
- Compute and interpret the mean, median, and mode.
- Identify when to use the range or standard deviation when describing a distribution's variability.
- Compute and interpret the standard deviation for a population or sample.
- Construct a scientific conclusion based on central tendency and variability statistics.

This chapter describes how to use graphs and measures of central tendency and variability to answer research questions. Specifically, you will learn about three types of graphs: frequency bar graphs, boxplots, and data plots. You will also learn three measures of central tendency: mean, median, and mode. Finally, you will learn two measures of variability: range and standard deviation. The graphs, measures of central tendency, and variability described in this chapter all illustrate how you can use data to answer a specific research question.

There is quite a bit of evidence that human touch is beneficial to our psychological and physical health. Massage is associated with reductions in anxiety, depression, and pain (Moyer et al., 2004), skin-to-skin contact can help infants gain weight (Boo & Jamli,

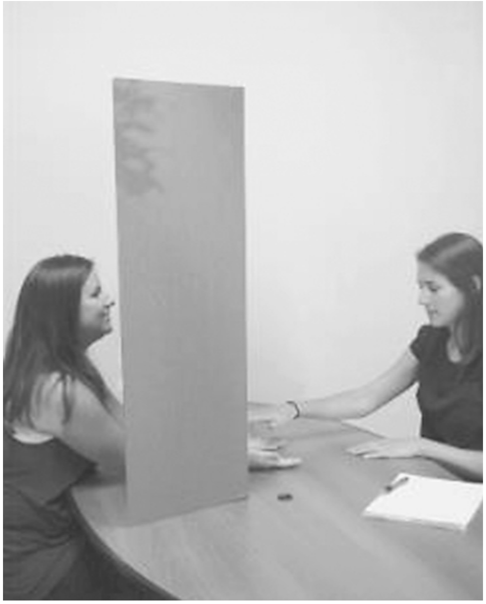


PHOTO 2.1: Representation of testing procedure

2007), and touch can improve immune system functioning (e.g., Field, 2010). Although there is little doubt of the benefits of physical touch, a treatment known as "therapeutic touch" (TT) is far more controversial. Therapeutic touch involves no actual physical contact. Instead, practitioners use their hands to move "human energy fields" (HEFs) in an attempt to help with relaxation, reduce pain, and improve the immune system (therapeutictouch.org). Can TT practitioners really sense HEFs?

Emily Rosa (who was just 9 years old at the time) and her colleagues (including her parents) investigated the basis of these TT claims by putting a sample of actual TT practitioners to the test (Rosa et al., 1998). As Photo 2.1 illustrates, individual practitioners sat at a table facing a large divider that prevented them from seeing their own hands or Emily. The practitioners placed both of their hands through the divider on the table, palms up. Practitioners were told to indicate whether Emily was holding her hand above their right or left hand. Emily began each trial by flipping a coin to determine where to place her hand. She then placed her hand 8 to 10 cm above one of the practitioner's hands. The practitioners had to "sense" the HEF allegedly emanating from Emily's hand to determine if Emily's hand was over their right hand or their left hand. Each practitioner went through a total of 10 of these trials. There is a link to a video about testing therapeutic touch on the textbook website.

FREQUENCY DISTRIBUTION GRAPHS AND TABLES

Bar Graphs

Before we examine the data Emily collected, we should consider what supportive evidence would look like. In other words, if TT practitioners can sense HEFs as they claim, they should be able to sense Emily's hand location correctly 10 out of 10 times. Even if we allow that sensing HEFs might be difficult for TT practitioners, the number of correct detections should be *close to* 10 out of 10, with most practitioners making 8, 9, or 10 correct detections. However, if they really cannot detect HEFs and the practitioners were really guessing, you would expect them to choose the correct hand location *an average* of 5 out of 10 times. Some practitioners may get lucky and guess more than 5 correct and others may get fewer than 5 correct, but the most common number correct would be about 5 of 10, *if the practitioners were guessing*. Figures 2.1a and 2.1b may help you visualize what supportive and disconfirming evidence would look like. These figures are frequency distributions. The height of the bar above each number of correct responses indicates how many practitioners gave that number of correct responses. So Figure 2.1a is an example of what supportive evidence would look like because the number of correct answers from the TT practitioners tend to "pile up" close to 10 out of 10 correct responses. Figure 2.1b shows an example of disconfirming evidence because TT practitioners' responses tend to pile up around 5 out of 10 correct responses, which is chance performance.

You should identify what supportive evidence would look like *before* you look at your data because doing so can help minimize the possibility that your personal preferences will influence your interpretation of the data. While scientists strive for objectivity, they are not immune from personal preferences. Therefore, they frequently follow bias-reducing procedures like identifying what constitutes supportive evidence before they look at the actual data.

Reading Question

1. On average, how many of the 10 trials should the therapeutic touch practitioners get right if they are just guessing?
   a. Eight.
   b. Five.
   c. Three.

Now that we have an idea of what to expect if TT practitioners can sense HEFs and what to expect if they cannot, we can look at the actual data from this study. The study collected data from 28 TT practitioners and the number of correct responses out of 10 for each practitioner was: 1, 2, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 6, 6, 7, 7, 7, and 8. You may recognize that presenting the data as a list of numbers is not very helpful. It certainly makes it difficult to draw a scientific conclusion. As mentioned in Chapter 1, it's a good idea to analyze the data in multiple ways. Doing so will likely help you create an accurate interpretation. Creating graphs of the data can be extremely helpful. Figure 2.2 displays a frequency bar graph of the raw data listed above. You worked with this type of graph in

FIGURE 2.1a ⬢ Example of What Supportive Evidence Might Look Like

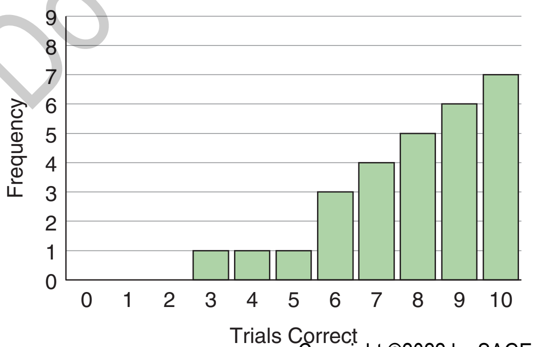


FIGURE 2.1b ⬢ Example of What Disconfirming Evidence Might Look Like

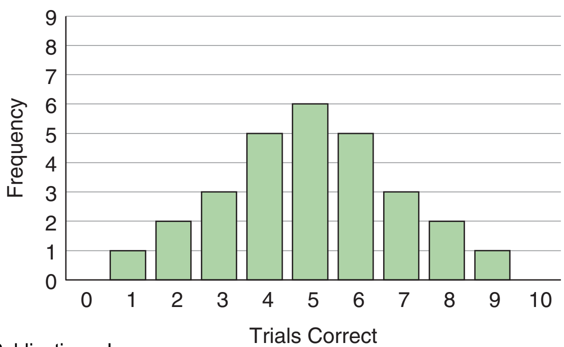
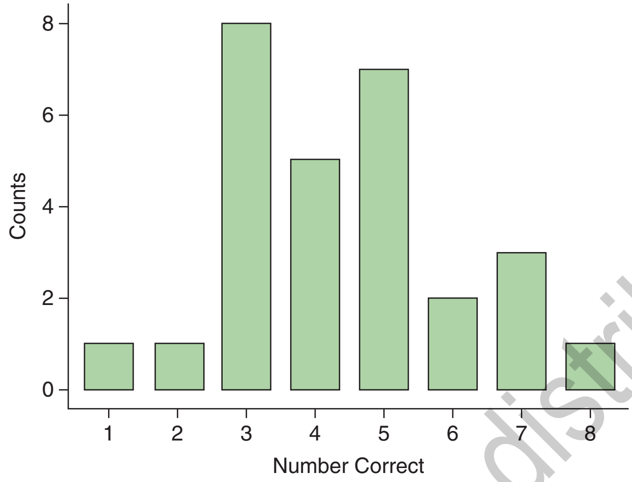


FIGURE 2.2 ● Frequency Bar Graph of Therapeutic Touch Data



the previous chapter. Organizing the data in this way conveys a lot of information. For example, it's much easier to see the "center" and "spread" of the data. The middle of the distribution is around 5 and there is variability with some practitioners doing better and worse than 5 to varying degrees. It also reveals that the most common score was 3 correct, with 8 individuals earning that score. Only 1 of 28 participants got 8 or more out of 10 correct.

Data Plots

Of course, there are other ways to graph these data. You might prefer some graph types more than others, but you should be able to read each of these graph types when you encounter it. Figure 2.3 displays the TT data as a "stacked" data plot. Data plots are similar to frequency distributions in that they plot all of the data points. The plot makes it clear that 1 practitioner got 8 correct, 3 got 7 correct, 2 got 6 correct, and so on, just as a frequency bar graph does. A stacked data plot conveys the same information as a frequency bar graph, but the information is just organized differently.

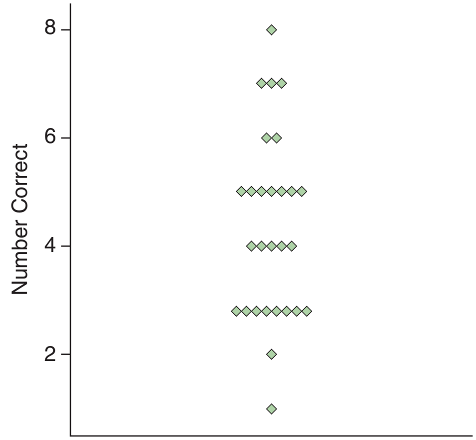
Reading Question

2. Bar graphs and stacked data plots both graphically display the
   a. averages.
   b. frequencies.
   c. medians.

Boxplots

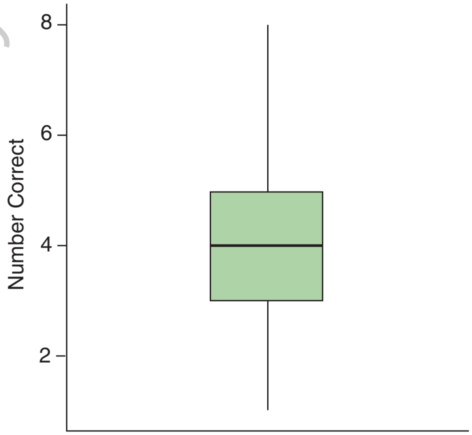
Figure 2.4 is a boxplot, a common graph type that summarizes the data using percentiles. In Chapter 1, you learned that a percentile tells you the percentage of scores that are at or below a particular value. For example, a score at the 25th percentile would mean that 25% of scores are at or below that score. The height of the bottom of the box in Figure 2.4 reveals the score

FIGURE 2.3 ● Stacked Data Plot of Therapeutic Touch Data



at the 25th percentile (i.e., 3); 25% percent of all the scores in the distribution are equal to or less than 3. The height of the horizontal line inside the box reveals the middle score, called the median (i.e., 4). Half of the scores in the distribution are above this value and half are below it. In other words, the height of this middle line reveals the 50th percentile, called the median. The top of the box reveals the 75th percentile (i.e., 5); 75% of all scores are equal to or less than this value. This particular statistic (the 75th percentile = 5) could help create our conclusion about TT. If 75% of all the TT practitioners are scoring at or below 5, which is chance performance, it suggests that they cannot detect HEFs. Finally, the vertical lines extending up and down from the box reveal the maximum and minimum scores. The top vertical line ends at 8, the maximum score, and the bottom vertical line ends at 1, the minimum score.

FIGURE 2.4 ● Boxplot of Therapeutic Touch Data



Reading Question

3. The middle line in a boxplot represents the (choose 2)
   a. 50th percentile.
   b. mean.
   c. median.
   d. mode.

Tables

Boxplots reveal the 25th, 50th, and 75th percentile scores. If you want additional percentile values, you need to look at a frequency distribution table. You worked with frequency tables

TABLE 2.1 ● Frequency Table of Therapeutic Touch Data

Levels	Counts	% of Total	Cumulative %
1	1	3.6	3.6
2	1	3.6	7.1
3	8	28.6	35.7
4	5	17.9	53.6
5	7	25.0	78.6
6	2	7.1	85.7
7	3	10.7	96.4
8	1	3.6	100.0

in Activity 1.1. Table 2.1 is a frequency table of the TT data. The first two columns of the table convey the same information as that found in frequency bar graphs and stacked data plots, namely, the number of people who got each score, but the last two columns convey additional information. The "% of Total" column reveals precisely what percentage of the scores were a specific value. For example, exactly 25% of the scores were 5's and 28.6% of the scores were 3's. The "Cumulative %" column conveys the percentage of values at and below a given score. For example, the frequency table reveals that 96.4% of the scores were below 8 out of 10 correct, the minimum value we identified as supportive evidence. We reasoned that if most of the TT practitioners identified 8, 9, or 10 hand locations correctly, it would suggest that TT practitioners can sense HEFs. The frequency table indicates that, rather than most meeting this criterion, only 1 or 3.6% reached this level of successful detection.

All these tables and graphs help us understand the TT data by displaying the center and spread of the data in one way or another. Whenever you are attempting to answer a research question by interpreting a distribution of data, you will want to describe the "center" and "spread" of the data. While graphing the data is helpful, it is rarely sufficient. Most of the time you will also want to compute additional statistics to help describe the center and spread of your data. Next, we describe the different statistics that describe the center of a distribution of scores and when you should use each.

Reading Question

4. The cumulative frequency column indicates the percent of scores _______________ a given value.
   a. at or below
   b. at or above
   c. greater than
   d. less than

CENTRAL TENDENCY: CHOOSING MEAN, MEDIAN, OR MODE

In Chapter 1, we recommended that you use multiple sources of evidence when constructing scientific conclusions. As mentioned above, graphs provide useful evidence about the center and spread of the data, but you will also want to compute specific values to describe the data more succinctly. When describing the center of a distribution you can choose from one of three options: the mean, the median, or the mode. The mean is the arithmetic average, the median is the score at the 50th percentile, and the mode is the most frequently occurring score. Each of these measures of central tendency defines the center differently; therefore, in some situations they provide dramatically different values for the center of a distribution, while in others they provide very similar values for the center. Each time you summarize your data, you must consider the variable's scale of measurement (i.e., nominal, ordinal, interval, or ratio) and the shape of the distribution to determine whether the mean, the median, or the mode offers the most accurate summary of the data.

When the data being summarized are nominal (i.e., when the data are categories rather than values), you must use the mode to summarize the center. For example, suppose your data is a list of students' academic majors. You asked students for their major and 17 responded psychology, 9 nursing, 7 sociology, and 8 social work. These academic majors are categories, not values. In other words, academic major is a nominal variable. With nominal data your only option is to count the number of responses in each category; the category with the highest count, the mode, is the center of the distribution. In this academic major data set, psychology is the mode with a count of 17.

Reading Question

5. When data are measured on a nominal scale, use the _______________ as the measure of central tendency.
   a. mean
   b. median
   c. mode

When the data being summarized are ordinal, you can use the mode or the median to describe the center of the data. For example, suppose we wanted to know the center class rank of students in a statistics course with 5 freshmen, 7 sophomores, 4 juniors, and 3 seniors. We could assign ordinal positions to each class rank so that freshman = 1, sophomore = 2, junior = 3, and senior = 4. This would result in the following distribution of class ranks: 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4. As mentioned, you have two options for describing the center of ordinal data—the median and/or the mode. The median is the value at the 50th percentile, which means that half of the scores are below this value and half are above it. *After the scores are arranged from lowest to highest,* the median is the middle value, or the one with the same number of scores above it and below it. For example, after the 19 class ranks are arranged from lowest to highest, the 10th-highest value would have nine scores above it and nine scores below it. The 10th highest score is a 2: 1, 1, 1, 1, 1, 2, 2, 2, 2, **2**, 2, 2, 3, 3, 3, 3, 4, 4, 4. Thus, the median value is sophomore. If a data set has an even number of scores, the median is the average of the middle two scores. You can also use the mode to represent the center of ordinal data but in most situations, you should use the median for ordinal data. In this case, the highest count, 7, is in the sophomore category so the mode of the data is sophomore. In this case, the median and the mode both say sophomore is the center of the distribution. As always, when multiple statistics provide the same answer, you should be more confident in that answer. When statistics differ, you should thoughtfully consider the situation to determine which is more accurate.

Reading Question

6. What is the median for this set of scores? 3, 3, 3, 4, 4, 5, 5, 5
   a. 3
   b. 4
   c. 5

Reading Question

7. In most situations, what measure of central tendency is used with ordinal data?
   a. mean
   b. median
   c. mode

When the data being summarized are interval or ratio, you can use the mode, median, or mean to describe the center of the data. For example, if you wanted to summarize how many texts you typically send/receive during an hour-long class, you could use the mean or the median because the number of texts sent/received in a class is ratio data. Suppose that during your last seven classes you sent/received the following numbers of messages: 7, 9, 38, 6, 7, 8, and 7. The mean number of text messages would be 11.7, but this is not really a typical value for you. During most classes (i.e., six of your last seven classes), the number of messages you sent was fewer than 11.7. The mean does not represent the center of these data very well because a single extreme score is inflating the mean, "pulling" it away from the center. Statisticians would consider the 38 value an **outlier** because it is *a very extreme score compared with the rest of the scores in the distribution*. The median is far less affected by extreme scores. The median would be 7 (6, 7, 7, **7**, 8, 9, 38), and this would be a much better summary of your in-class texting habits. The mode of these data would also be 7. In this case both the median and mode say 7 is the center and the mean is "pulled away" from center by the outlier 38. Therefore, the best summary of center for these data is 7. Table 2.2 summarizes when to use each measure of central tendency. In general, when the data are interval or ratio and the data are symmetrically shaped (e.g., similar to Figure 2.1b, not 2.1a) and free from outliers, the mean will be the best measure of center. The TT data will illustrate this point below.

TABLE 2.2 ⬢ When to Use Measures of Central Tendency

Type of Data	Measure of Central Tendency to Use
Nominal	Mode
Ordinal	Median
Interval or Ratio	Mean Additional considerations: If data are highly skewed or there are extreme outliers, use the median.

Reading Question

8. By adding all the scores and then dividing by the number of scores you get the
   a. mean.
   b. median.

Reading Question

9. The ________________ has half of the scores above it and half of the scores below it.
   a. mean
   b. median

Reading Question

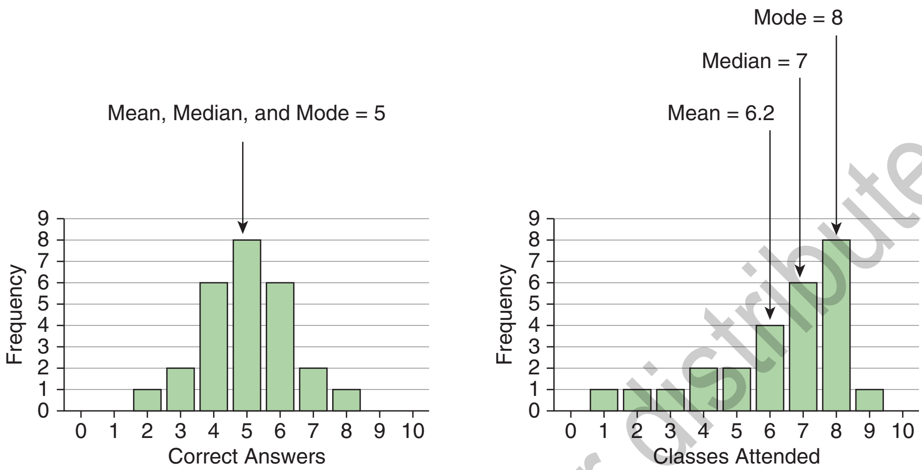
10. When the data are interval and there are extreme scores in the distribution, you should use the ________________ to describe the center.
    a. mean
    b. median
    c. mode

Reading Question

11. Extreme scores are also called
    a. outliers.
    b. modes scores.

As mentioned, when the data are interval or ratio, you can use the mean, median, or mode; you should thoughtfully consider which provides the best description of center. The shape of the distribution should help you make this determination. The median will likely be the better choice when the distribution is asymmetrical or skewed. A distribution of scores is skewed if the "tail" on one side of the distribution is substantially longer than the tail on the other side. The bar graphs in Figure 2.5 illustrate a symmetrical distribution and a negatively skewed distribution, respectively. Generally, the mean is better for symmetrical distributions, and the median is better for highly skewed

FIGURE 2.5 ⬢ A Bar Graph of Symmetrical Data and a Bar Graph of Negatively Skewed Data



(i.e., asymmetrical) distributions. In asymmetrical distributions, the mean is "pulled away" from the center toward the distribution's longer tail. This fact is illustrated by the asymmetrical graph in Figure 2.5. It is worth noting that when a distribution is symmetrical (or close to being symmetrical), the mean, the median, and the mode are all very similar in value. You can see this in Figure 2.5.

Reading Question

12. When interval or ratio data is highly skewed, use the _____ to describe the center.
    a. mean
    b. median
    c. mode

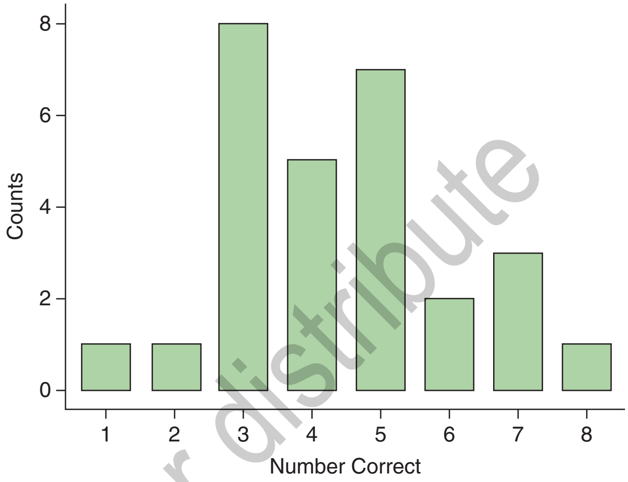
Reading Question

13. When a distribution of scores is skewed, the median and the mean will be similar.
    a. True
    b. False

COMPUTING MEASURES OF CENTRAL TENDENCY

Now that you know when each measure of central tendency can be used, let's apply this knowledge to the TT data. The number of correct detections is ratio data, therefore the mean, median, and mode are all viable measures of the data's center. The first step is to look at the data to determine if there are outliers or if the distribution is highly skewed. We can use the frequency bar graph we created above for this; Figure 2.6 reproduces this graph. Glancing at the frequency distribution suggests there are no outliers and that the data are close to symmetrical. Therefore, the mean, median, and mode are all options for summarizing the data's central tendency. In situations like this, compute all three, then thoughtfully consider which of them provides the more accurate representation of the distribution's center. You compute the mean by adding all the scores (ΣX) and dividing by the number of scores (N). The frequency distribution indicates that there is one person with a score of 1, one person with a score of 2, eight with a score of 3, and so on. To obtain the sum of all the scores (ΣX), you could add each individual number, but it is more efficient to work with the data as they are presented in the graph. So you can multiply each score by the count of people who had that score. This is illustrated below:

FIGURE 2.6 ● Frequency Bar Graph of Therapeutic Touch Data



$$M = \frac{\sum X}{N} = \frac{1(1)+2(1)+3(8)+4(5)+5(7)+6(2)+7(3)+8(1)}{28} = \frac{123}{28} = 4.39$$

In this text we use the symbol M to represent the mean of a sample. When working with a population, we use the Greek letter μ (pronounced "myoo"). More generally, you will find that researchers use Greek letters to represent population parameters and Roman letters to represent samples.

Reading Question

14. The Greek letter μ represents the _______________ mean.
   a. sample
   b. population

To find the median you need to determine how many scores are in the distribution. In this case there are 28, so $N = 28$. When N, the number of scores, is even you find the middle of the middle two scores. So, for the TT data with 28 scores the middle two scores are the 14th and 15th scores (remember that the scores must be arranged in ascending order). Using the frequency bar graph, you can determine that the 1st score is 1, the 2nd score is 2, the 3rd through 10th are 3, and the 11th through 15th are 4. Therefore, the middle between the 14th number which is 4 and the 15th number which is also 4 would be the median of 4.

To find the mode you simply look at the frequency bar graph and find the most common score; in this case it is 3.

Now that you have all three measures of central tendency for the TT data you need to determine which offers the best measure of center. As noted above, when the distribution is close to symmetrical, the mean, median, and mode will be similar. In this case they were 4.39, 4, and 3, respectively. These values are similar, but the mean is the best measure in this case. As noted above, when the data are interval or ratio the mean will probably be the best measure as long as there are no outliers and the distribution is close to symmetrical.

In the next section we describe two measures of data spread (i.e., variability), when you should use each, and how to compute them.

Reading Question

15. The most common score in a distribution is called the
    a. median.
    b. mode.

Reading Question

16. When a distribution of scores is symmetrical, the mean is _____________ the median.
    a. very close to
    b. higher than
    c. lower than

VARIABILITY: RANGE OR STANDARD DEVIATION

Now that you know the mean number of correct detections by the TT practitioners was 4.39 (M = 4.39), you also need to describe the spread or variability of the scores. Inspecting the frequency bar graph tells you that there is variability; not everyone got 4.39 correct answers. Now you want to compute a statistic that summarizes the amount of variability in the distribution. There are two choices: the range and the standard deviation. The **range** is the easiest measure of variability, *the difference between the highest and lowest scores*. For the TT data the range is 8 − 1 = 7.[1] The range is a poor measure of variability because it is very insensitive. By insensitive, we mean the range is unaffected by any of the middle scores—it ignores them entirely. The middle scores could all be different or identical and the range would not change as long as the highest score (i.e., 8) and the lowest score (i.e., 1) do not change. However, when working with ordinal data, the range is the best option available. When working with interval/ratio data, you can use a more sensitive measure of variability that considers

[1]Some software computes the range as the highest value – lowest value + 1.

all the scores in the distribution, and not just the highest and the lowest. When the data being summarized are interval or ratio, the most common measure of variability is the **standard deviation**. *The standard deviation tells you the typical, or standard, distance each score is from the mean.*[2] In the next section we describe how to compute and interpret the standard deviation.

Reading Question

17. Why is the range a poor measure of variability?
    a. It uses only two values rather than all the values in the distribution.
    b. It is overly sensitive to changes in the middle of the data.

Reading Question

18. What characteristic of a distribution of scores does a standard deviation describe?
    a. How far scores are from the mean.
    b. How spread out the scores are.
    c. The variability of scores in a distribution.
    d. All of the above.

Reading Question

19. What measure of variability should be used when the data are interval/ratio?
    a. Standard deviation.
    b. Range.

STEPS IN COMPUTING A POPULATION'S STANDARD DEVIATION

When describing the variability of data, you need to know if the data are from everyone in the population or from a sample of the population. If the TT data came from all the TT practitioners in Reno, Nevada, you might consider this the entire population. If you limit your conclusions to only those practitioners in Reno, this would be completely appropriate. If, on the other hand, the data are only a subset of the TT practitioners in Reno, Nevada, the data would be considered a sample. Computing the standard deviation of a sample is done slightly differently from computing it for a population. We describe how to compute both starting with the procedure for a population.

[2] It is tempting to describe the standard deviation as the average deviation from the mean, but this is not technically correct because the deviation scores always sum to zero and so the average deviation is always zero. A different measure of variability, the mean absolute deviation, computes the average of each score's absolute deviation from the mean (i.e., it ignores the sign of the deviation). However, the mean absolute deviation is rarely used. You will sometimes hear people talk about the standard deviation as the average deviation. Although this is not technically accurate, thinking about the standard deviation as the average deviation is fine.

Steps 1–3: Computing the Sum of the Squared Deviation Scores (*SS*)

Computing the standard deviation consists of five steps. Steps 1–3 are shown in Table 2.3. Focus on understanding what you are doing at each step rather than simply doing the calculations. The scores (*X*s) from the TT practitioner study are in the first column of Table 2.3.

The standard deviation measures the standard (or typical) distance each score is from the mean. Thus, to compute the standard deviation, you first need to determine how far each score is from the mean, called a *deviation score*. You compute each score's distance from the population mean by $X - \mu$, where X is the score and μ is the mean of the population. For example, in the TT data the *possible* scores ranged between 0/10 correct to 10/10 correct. The actual scores ranged between 1/10 to 8/10, with some of these values being produced by more than one practitioner. Table 2.3 displays each of the actual scores and how each score is converted to a deviation score by subtracting the population mean ($\mu = 4.392857143$).

This is done in the second column of Table 2.3. We will not typically compute answers with so many decimal points but to get the same value as your computer you will need your computations to be precise, specifically, one more decimal place than you want your final answer to have. If you want an answer with two decimal places you must compute everything to three places. The second step is to square each deviation score, as shown in the third column of Table 2.3. The third step is finding the sum of the squared deviation scores, called the SS or sum of squares. This is completed at the bottom of the third column in Table 2.3.

As you can see, computing the *SS* with the $SS = \sum (X - \mu)^2$ formula can be tedious if the data set is large and/or the mean of the scores is not a whole number. There is an alternative way to compute the *SS* called the computational method. It will give you the same answer as the method illustrated in Table 2.3 (the definitional method). The computation method formula is

$$SS = \sum X^2 - \frac{(\sum X)^2}{N}$$

To find the $\sum X^2$ you would first square each X and then sum the squared values. To find the $(\sum X)^2$ you would sum the X's and then square that value. N is the number of scores. For this example,

$$SS = \sum X^2 - \frac{(\sum X)^2}{N} = 615 - \frac{123^2}{28} = 74.67857$$

TABLE 2.3 ⬢ Computing the Sum of the Squared Deviation Scores

	Step 1	Step 2
Score (X)	Find Each Deviation Score ($X - \mu$)	Square Each Deviation Score $(X - \mu)^2$
1	1 – 4.392857143 = –3.39286	11.51148
2	2 – 4.392857143 = –2.39286	5.725765
3	3 – 4.392857143 = –1.39286	1.940051
3	3 – 4.392857143 = –1.39286	1.940051
3	3 – 4.392857143 = –1.39286	1.940051
3	3 – 4.392857143 = –1.39286	1.940051
3	3 – 4.392857143 = –1.39286	1.940051
3	3 – 4.392857143 = –1.39286	1.940051
3	3 – 4.392857143 = –1.39286	1.940051
3	3 – 4.392857143 = –1.39286	1.940051
4	4 – 4.392857143 = –0.39286	0.154337
4	4 – 4.392857143 = –0.39286	0.154337
4	4 – 4.392857143 = –0.39286	0.154337
4	4 – 4.392857143 = –0.39286	0.154337
4	4 – 4.392857143 = –0.39286	0.154337
5	5 – 4.392857143 = 0.607143	0.368622
5	5 – 4.392857143 = 0.607143	0.368622
5	5 – 4.392857143 = 0.607143	0.368622
5	5 – 4.392857143 = 0.607143	0.368622
5	5 – 4.392857143 = 0.607143	0.368622
5	5 – 4.392857143 = 0.607143	0.368622
5	5 – 4.392857143 = 0.607143	0.368622
6	6 – 4.392857143 = 1.607143	2.582908
6	6 – 4.392857143 = 1.607143	2.582908
7	7 – 4.392857143 = 2.607143	6.797194

7	7 – 4.392857143 = 2.607143	6.797194
7	7 – 4.392857143 = 2.607143	6.797194
8	8 – 4.392857143 = 3.607143	13.01148
$M = \sum X / N$ = 123/ 28 = 4.392857143		**Step 3** **Sum the Deviation Scores** $SS = \sum (X - \mu)^2$ = 74.67857

Note that this *SS* value is identical to that found with the definitional method in Table 2.3.

Reading Question

20. A deviation score measures
    a. the typical distance all the scores are from the mean.
    b. the distance of an individual score from the mean.

Reading Question

21. *SS* stands for the
    a. standard deviation.
    b. sum of the squared deviation scores.
    c. sum of the deviation scores.

Step 4: Computing the Variance (σ^2)

Again, your goal is to compute the typical, or standard, deviation of the scores from the mean of the TT data. We cannot compute the average deviation score because their sum is always zero. So, instead, we compute the average *squared* deviation score, which is called the **variance** (σ^2, lowercase sigma squared). When computing any average, we divide the sum of values by the number of values. Therefore, in this case we divide the sum of the squared deviation scores (i.e., the *SS*) by the number of squared deviations (i.e., *N*). The result is the mean of the squared deviation scores, the variance, σ^2. Therefore,

$$\text{Population variance: } \sigma^2 = \frac{SS}{N} = \frac{74.67857}{28} = 2.667.$$

Reading Question

22. The variance (σ^2) is the
    a. typical squared deviation from the mean.
    b. typical deviation from the mean.

Step 5: Computing the Standard Deviation (σ)

You *squared* the deviation scores before you summed them and then divided the sum by *N* to get the variance. This means that the variance is the typical *squared* deviation of all the scores from the mean. While informative, this value is not very intuitive to think about. It is much easier to think about the typical deviation of scores from the mean. Therefore, you convert the typical *squared* deviation into the typical deviation by taking the square root of the variance. The square root of the variance is the typical or standard deviation of scores from the mean:

$$\text{Population standard deviation: } \sigma = \sqrt{\sigma^2} = \sqrt{\frac{SS}{N}} = \sqrt{\frac{74.67857}{28}} = \sqrt{2.667} = 1.633.$$

In this population, the standard deviation is 1.63. Some scores are more than 1.63 away from the mean of M = 4.39 and other scores are less than 1.63 away from the mean, but the "typical" distance of all the scores is σ = 1.63. For example, in this distribution the scores of 3, 4, 5, and 6 are all closer to the mean of M = 4.39 than the standard deviation of σ = 1.63. Other scores in the distribution are farther from the mean M = 4.39 than the standard deviation of σ = 1.63, specifically scores of 1, 2, 7, and 8. It will always be the case that some scores deviate more than the σ and some deviate less, but the σ is the standard or typical amount of deviation from the mean of the distribution of scores.

Reading Question

23. The standard deviation (σ) is
    a. how far all the scores are from the mean.
    b. the *typical* distance all the scores are from the mean; some scores will be farther away and some closer, but this is the typical distance from the mean.

Table 2.4 displays the five steps to computing the standard deviation of a population. Familiarize yourself with the verbal labels as well as their symbolic equivalents because you will be using both in future chapters. You should notice that there are two SS formulas. While these formulas are mathematically equivalent (meaning they yield the same answer), the second formula is much easier to use with computing values for larger data sets; therefore, it is called the computational formula. However, most of the time you will use statistical software to compute the standard deviation of your data, so it is more important that you understand what the standard deviation measures than how to compute it by hand for large data sets. The upshot is, you need to know that the standard deviation is a measure of variability in the data; it is the *standard* amount of variability from the mean; some scores vary more from the mean and some vary less.

Reading Question

24. What symbol represents the standard deviation of a population?
    a. SS.
    b. σ.
    c. σ^2.

Reading Question

25. Which equation defines the sum of the squared deviation scores?

a. $\sum(X-\mu)^2$

b. $\sum(X-\mu)$

c. $\sqrt{\sigma^2}$

Reading Question

26. Which equation(s) yield the *SS*? Select all that apply.

a. $\sum(X-\mu)^2$

b. $\sum(X-\mu)$

c. $\sum X^2 - \frac{(\sum X)^2}{N}$

STEPS IN COMPUTING A SAMPLE'S STANDARD DEVIATION

You just computed the standard deviation of the TT data assuming you had data from all the TT practitioners in a population, say a given city like Reno, Nevada. However, most of the time you will have data from a *sample* of the population rather than an entire population. This changes how you compute the standard deviation slightly.

Steps 1–3: Obtaining the *SS*

Computing the sum of the squared deviation scores (*SS*) is identical for a sample and population. The *SS* for the TT data is *SS* = 74.68.

Reading Question

27. The *SS* is computed in the same way for a sample and a population.

a. True

b. False

Step 4: Computing the Sample Variance (SD^2)

The first computational difference arises when computing the sample variance, SD^2. When computing the *population variance,* you divided the *SS* by *N*. However, when computing the *sample variance,* you divide the SS by $N - 1$. This is the only difference when computing the sample variance, SD^2:

$$SD^2 = \frac{SS}{N-1} = \frac{74.68}{28-1} = 2.77.$$

Why do you divide by $N - 1$ for sample variances and N for population variances? This is a somewhat complicated issue. A simple explanation is that samples are less variable than populations, and without the $N - 1$ adjustment, our sample variability would not estimate the population's variability as well—it would be too low. For example, the variability of a sample of TT practitioners' data will be less than the entire population. More data tend to create more variability. The difference in variability between smaller samples and larger populations is a serious problem if you are trying to use a sample to estimate a population's standard deviation. So you need to do a computational adjustment when using a sample to estimate a population's variability; if you don't, your variability estimate will tend to be too low. The computational adjustment statisticians determined to be most accurate in most situations is dividing the SS by $N - 1$ rather than by N. You will typically compute the standard deviation using statistical software and should be aware that software is almost always computing the sample standard deviation and not the population standard deviation. If only one standard deviation is provided, you should assume that it is the sample standard deviation.

Reading Question

28. When using a sample to estimate a population's variability, the SS is divided by $N - 1$ rather than by N to correct for a sample's tendency to
    a. overestimate the variability of a population.
    b. underestimate the variability of a population.

Step 5: Computing the Sample Standard Deviation (*SD*)

Take the square root of the sample variance:

$$SD = \sqrt{SD^2} = \sqrt{2.766} = 1.663.$$

The verbal labels corresponding to each computational step for a sample's standard deviation are identical to those used when computing a population's standard deviation. However, the sample's symbolic equivalents are Roman letters rather than Greek letters (Table 2.4).

Reading Question

29. What symbol represents the standard deviation of a sample?
    a. SD
    b. SD^2
    c. SS

TABLE 2.4 ● Summary of Five Steps to Computing Sample and Population Standard Deviation

Step	Verbal Label	Symbolic Equivalent		Equation	
		Population	Sample	Population	Sample
1	Deviation score			$(X-\mu)$	$(X-M)$
2	Square the deviation scores			$(X-\mu)^2$	$(X-M)^2$
3	Sum of squared deviation scores	SS	SS	Definitional: $SS=\Sigma(X-M)^2$ Computational: $SS=\sum X^2-\frac{(\Sigma X)^2}{N}$	
4	Variance	σ^2	SD^2	$\sigma^2=\frac{SS}{N}$	$SD^2=\frac{SS}{N-1}$
5	Standard deviation	σ	SD	$\sigma=\sqrt{\frac{SS}{N}}$	$SD=\sqrt{\frac{SS}{N-1}}$

Reading Question

30. When *estimating* the variance of a population from a sample, you are performing ________________ so divide the SS by ________________.
    a. descriptive statistics; N.
    b. inferential statistics; $N-1$.

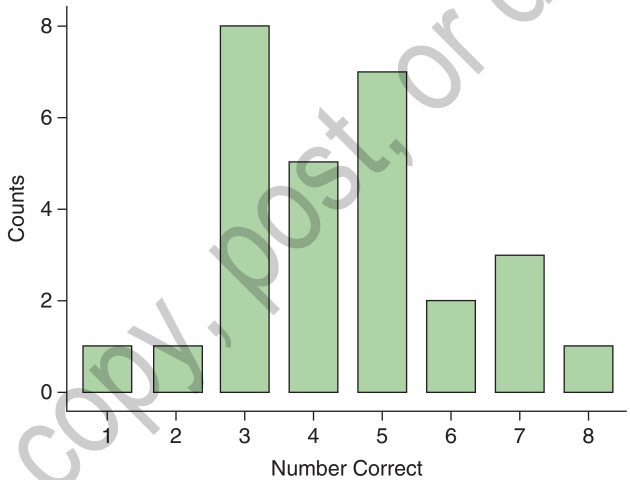
CONSTRUCTING A SCIENTIFIC CONCLUSION

Now that you have gathered multiple sources of statistical evidence including graphs of the data, several measures of central tendency (i.e., mean, median, and mode), and a measure of variability (i.e., the sample standard deviation), you must construct a scientific conclusion. Based on all this evidence, what do the data suggest? Can TT practitioners sense HEFs?

The frequency bar graph suggests that the data's distribution is centered close to chance performance (i.e., 5 correct out of 10 trials) rather than the scores of 8, 9, or 10 correct responses (i.e., the values determined before collecting data to represent supportive evidence). The boxplot reveals that 75% of TT practitioners gave 5 or fewer correct responses. Additionally, the frequency table indicates that only 1 out of 28 practitioners reached the predetermined level of supportive evidence (8, 9, or 10 correct responses). Supportive evidence would have most of the TT practitioners obtaining 8 or higher correct responses, and not just the 1 observed in this study. The empirical evidence from Rosa et al. (1998) suggests that TT practitioners cannot detect HEFs. However, before constructing your scientific conclusion about TT practitioners, you should consider the evidence provided by other studies in the scientific literature. If other studies found similarly negative evidence pertaining to TT, your scientific conclusion could be stated more confidently. Conversely, if the scientific literature offers a lot of supportive evidence for TT, you should consider why your study's results seem to deviate from the results of others. If you reviewed the literature on TT, you would find a mixed bag. The more scientifically rigorous studies that incorporate well-designed control groups report highly negative evidence (i.e., evidence suggesting TT does not help with relaxation, pain, or the immune system) (e.g., Frank et al., 2007; Johnston et al., 2013). However, TT does have its proponents. Generally, the evidence in favor of TT comes from studies that are less well designed, suggesting their conclusions might not be as valid as the more-rigorous studies. In addition to weighing the evidence in the scientific literature, you should consider additional information as well. You might ponder, "What is a HEF?" Is there *any* way to reliably measure HEFs? Is there evidence that HEFs cause health problems? If HEFs were real, an assumption at this point, how might TT practitioners "sense and move" them? Can the physiological mechanism by which TT practitioners purportedly sense and move HEFs be explained by our existing knowledge of human biology or do explanations of TT effectiveness rely on "alternative" or "mystical" explanations? If the mechanism underlying a phenomenon cannot be explained, it is wise to be skeptical. Where science is concerned, extraordinary claims require extraordinary evidence. Now, based on all the available evidence, what is the most prudent scientific conclusion pertaining to TT? The following paragraph represents a summary of the TT data collected by Rosa et al. (1998) pertaining to the question, "Can TT practitioners sense HEFs?" The paragraph first describes the empirical evidence and then places the evidence in a broader context by considering the evidence in the scientific literature as well as the additional questions that would need to be answered before offering a scientific explanation of TT. An example of a scientific conclusion follows:

> When TT practitioners had to sense a HEF allegedly emanating from an experimenter's hand that was held above their left or right hand, their mean number of correct responses (M = 4.39, SD = 1.66) was lower than 5 correct out of 10 trials, *below* chance performance. The mean number of correct responses and the frequency

bar graph of individual participants' correct responses (shown below) both suggest that TT practitioners performed as if they were guessing the hand location. Furthermore, the data show that only 1 out of 28 practitioners generated 8 or more correct responses across 10 trials, with 8 being the agreed-on level of supportive evidence. If TT practitioners could sense HEFs, one would expect most of them to be correct 8 or more times. However, the data from this current study suggest that TT practitioners were guessing. Further supporting this conclusion, the previous research on TT that used similarly rigorous control procedures also found no evidence to support the claims of TT practitioners (e.g., Frank et al., 2007; Johnston et al., 2013). Finally, the physiological mechanisms by which HEFs might be detected and moved by TT practitioners remain unspecified, making the acceptance of TT as a scientifically supported therapy highly dubious. Based on the available data, it seems likely that TT offers little more in the way of benefits than placebo effects.



Reading Question

31. Which of the following is NOT in the above scientific conclusion?
   a. The mean number of correct responses was $M = 4.39$.
   b. The standard deviation of the correct responses was $SD = 1.66$.
   c. The mean value was less than that expected by chance.
   d. Other studies on TT that used rigorous methods also found results that challenged TT claims.
   e. Currently there are no scientific theories that can explain how TT is supposed to work.
   f. All of these are in the above scientific conclusion.

Activity 2.1

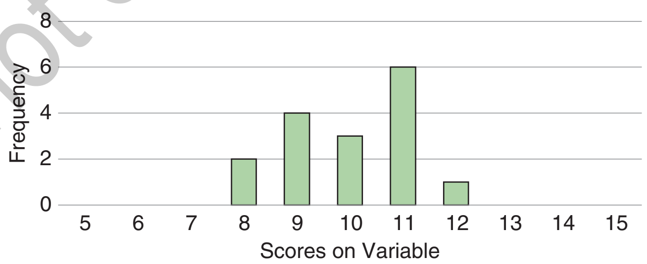
Computing Central Tendency and Standard Deviation

In the previous activity you used Milgram's (1974) data, in the form of frequency distributions, to draw scientific conclusions about obedience to authority. In most research situations, it will help you construct better scientific conclusions if you can describe the center and variability of a distribution of scores. The "center" of a distribution reveals the "average" or "typical" response of participants, which is frequently of interest. Equally interesting, the amount of variability in a distribution can reveal if participants' responses tended to be very similar to each other or if individuals provided very diverse responses. This activity will help you interpret a distribution's center and spread.

1. When doing research, it is important to
   a. understand the center of a distribution of scores.
   b. understand the variability of a distribution of scores.
   c. understand both the center and the variability of a distribution of scores.

Understanding the Mean as the Balancing Point

The mean, median, and mode all measure the center of a distribution of scores. However, each of these statistics defines "center" differently. The **mode** defines the center as *the most common score*. The **median** defines the center as *the middle score*. The **mean** defines the center in a more sophisticated way, specifically as *the value that balances positive and negative deviation scores*. In this activity, we use frequency histograms to help you understand how the mean defines "center." The data in the graph below are scores from a population.



2. The mode of this distribution is _______________ .
3. The median of this distribution is _______________ .
4. How many scores are in this distribution? (i.e., What is N?) _______________

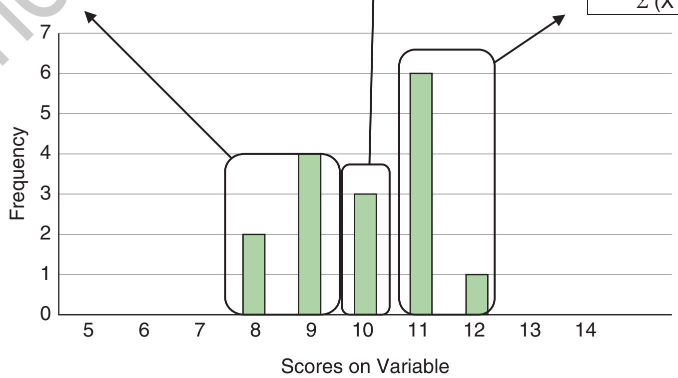
5. Compute the mean of this distribution by adding up all the scores ($\sum X$) and dividing by the number of scores (N). $M = \frac{\sum X}{N}$

6. Although you have probably computed the mean many times, you probably don't understand how it "defines" center. The mean defines center by balancing **deviation scores**. In other words, *the mean is the only value that perfectly balances all the positive and negative deviation scores in the distribution.* What is a deviation score? Every value in a distribution is a certain distance from the mean; this distance from the mean is the value's **deviation score.** In this case you are working with a sample so a deviation score is computed as a score (X) minus the mean (M), or ($X - M$). For a population, the computations are the same but the notation for the mean is μ rather than M, therefore the deviation score notation when dealing with a population is ($X - \mu$). Computing deviation scores will help you understand how the mean balances them. Use the formula ($X - M$) on each of the scores in the distribution below. For example, both scores of 8 have a deviation score of ($X - M$) = (8 – 10) = –2. The graph below has three boxes: one box for the six scores that are less than the mean, a second for the three scores at the mean, and a third for the seven scores that are greater than the mean. Compute the deviation scores for each score and put them in the tables below. After you have computed the deviation scores, sum the deviations ($\sum(X - M)$) that are above, below, and at the mean.

Scores Below the Mean	
Score	Deviation (X – M)
8	
8	
9	
9	
9	
9	
Σ (X – M) =	

Scores At the Mean	
Score	Deviation (X – M)
10	
10	
10	
Σ (X – M) =	

Scores Above the Mean	
Score	Deviation (X – M)
11	
11	
11	
11	
11	
11	
12	
Σ (X – M) =	



7. What is the sum of the positive deviation scores?
8. What is the sum of the negative deviation scores?
9. Notice how the sum of all the positive deviations and the sum of all the negative deviations balance each other. The mean is the only value that perfectly balances all of the deviation scores. Therefore, the sum of all the deviation scores, $\sum (X - M)$, will ALWAYS equal what value exactly? _______________
10. Which of the following statements about the mean is true?
    a. The mean is ALWAYS the exact point at which the sum of positive and negative deviation scores balance each other.
    b. The number of deviation scores above the mean is ALWAYS equal to the number of deviation scores below the mean.
    c. The mean is ALWAYS the score with the highest frequency.

Computing the Standard Deviation

When scores are measured on an interval or ratio scale of measurement, the mean is the most common measure of a distribution's "center" because its balancing of deviation scores provides a more precise measurement than the mode or median. In addition to describing a distribution's center, you also need a precise measure of a distribution's variability, or how much scores tend to deviate from the center. When the scores are measured on an interval or ratio scale, the standard deviation is the most used statistic. The **standard deviation** defines variability as *the typical deviation from the mean of the distribution*. Just as deviation scores helped you understand how the mean defines center, deviation scores also describe how the standard deviation defines variability.

11. The mean is a measure of center and the _______________ is a measure of variability. Both are based on _______________ scores.
12. Because deviation scores always sum to zero, we cannot simply use the average deviation score as a measure of variability. To address this problem, square each deviation score in Activity Table 1 that follows. Then sum the squared deviation scores to create a statistic called the sum of the squared deviation scores (*SS*). This statistic is an intermediate step to computing the standard deviation.
13. This table lists every score individually so you can see how to compute the *SS*, but you can compute it more efficiently by using the frequencies. Use Activity Table 2 that follows to compute the *SS* using this more-efficient table. This data set is small and so you might not notice a significant time savings, but you would prefer this

method if you were working with a very large data set. If you do this correctly, you will get the same answer as you did in the previous question.

ACTIVITY TABLE 1

Score	Deviation Scores: $(X - M)$	Squared Deviation Scores: $(X - M)^2$
8	8 − 10 = −2	
8	8 − 10 = −2	
9	9 − 10 = −1	
9	9 − 10 = −1	
9	9 − 10 = −1	
9	9 − 10 = −1	
10	10 − 10 = 0	
10	10 − 10 = 0	
10	10 − 10 = 0	
11	11 − 10 = 1	
11	11 − 10 = 1	
11	11 − 10 = 1	
11	11 − 10 = 1	
11	11 − 10 = 1	
11	11 − 10 = 1	
12	12 − 10 = 2	
		SS =

ACTIVITY TABLE 2

Score (X)	Count (f)	Step 1: Deviation Score $(X - \mu)$	Step 2: Square Deviation Scores $(X - \mu)^2$	Step 3: Multiply $f * (X - \mu)^2$, then sum
8	2			
9	4			
10	3			

11	6			
12	1			
				SS =

15. The sum of the squared deviation scores (*SS*) is a measure of variability, but not one that is easily interpreted. The next step is to divide the *SS* by number of scores (*N*). This gives you the variance, which is the mean squared distance from the mean.

$$\sigma^2 = \frac{SS}{N}$$

16. We squared the deviation scores when computing the *SS*. To put the values back into original units, take the square root of variance to find the standard deviation.

$$\sigma = \sqrt{\frac{SS}{N}}$$

17. In this example, the data came from a population. If we were using sample data, all of the calculations above would be the same EXCEPT you would divide the *SS* by $N - 1$ rather than *N*. Compute the standard deviation as if we were using sample data.

$$SD = \sqrt{\frac{SS}{N-1}}$$

More Practice With Computations

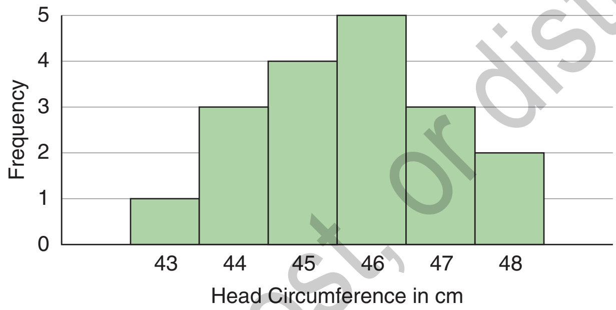
18. We will rely heavily on software to compute the mean and standard deviation. However, it can be helpful to work through the computations by hand with small data sets to help you understand what the mean and standard deviation are measuring. Compute the mean and standard deviation for these four numbers from a sample: 68, 61, 72, 70. You should find that the mean is 67.75 and the *SD* is 4.79.

19. Compute the mean and standard deviation for these five numbers from a population: 5, 6, 3, 2, 7.

Activity 2.2

Identifying Causes of Variability

A developmental psychologist studying how the neonatal environment affects infant development needs to develop a reliable way to collect the physical measurements of newborn infants. He knows that, at every doctor's visit, the child's height, weight, and head circumference are measured, typically by nurses; he wants to know how accurate their measurement procedures are. After the nurses agree to participate, each nurse measures the head circumference of a very realistic doll of a 1-year-old infant in centimeters. The 18 nurses' measurements are graphed below.



1. Given that all nurses measured the head of the same doll, all 18 head circumference measurements should be the same (i.e., there should be no variability in the measurements). However, the above graph clearly illustrates that there was variability. Why was there variability in the measurements of the doll's head circumference? (Select all that apply.)
   a. Some nurses held the tape measure tighter around the infant's head while others held it looser.
   b. Each nurse put the tape measure in a slightly different place on the doll's head.
   c. Some nurses misread the tape measure.
   d. The doll's head changed size between measurements.
2. In the above question, all the variability in scores was created by **measurement error** because everyone was measuring the same thing and, therefore, should have obtained the same score. Unfortunately, measurement error is always present. No matter what you measure, you can never measure it perfectly every time. You can, however, reduce the amount of measurement error. In the context

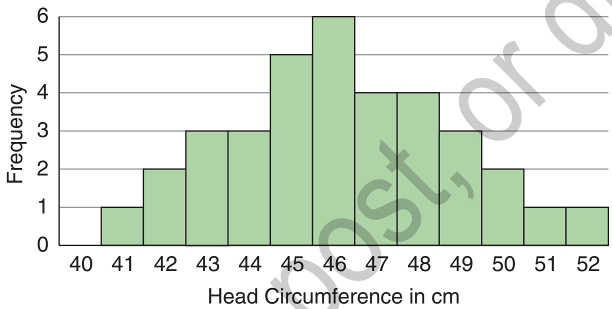
of measuring an infant's head circumference, how could the developmental psychologist and/or nurses reduce the variability in scores created by measurement error (i.e., what could they do to increase the accuracy/reliability of each measurement)? Select all that apply.

a. Give the nurses a lot of practice measuring different dolls' heads.

b. Train the nurses to use a consistent degree of tension in the tape measure.

c. Use a tape measure that records only centimeters, and not inches.

3. The following week each nurse measured the head circumference of one or more infants. The head circumference measurements for 35 *different* infants are graphed below:



Did measurement error create some of the variability in the scores graphed above?

a. No, there is no measurement error. The head circumferences are only different because different infants were measured.

b. Yes, there is measurement error. There is always some potential for measurement error any time a measurement is taken.

4. In addition to measurement error, something else is also creating variability in the distribution of 35 scores (i.e., head circumferences). Besides measurement error, what is another reason for the variability in the above distribution of 35 infants' head circumferences?

a. Treatment variability; a treatment created this variability.

b. Individual differences; some infants have larger heads than others.

5. In the previous question, the variability in head circumferences was created by both *measurement error*, which is always present, and the fact that the 35

infants' heads varied in size, called **individual differences** variability. In which of the following distributions of scores would there be *more* variability created by individual differences?

a. The head circumferences of 50 first graders.

b. The head circumferences of 50 elementary school children (first through fifth graders).

6. In many research situations, researchers *generate* variability with **different treatment conditions**. For example, if you thought physically touching prematurely born infants would increase their growth, you could conduct a study with two samples of prematurely born infants. Infants in Group 1 could be touched with skin-to-skin contact for 6 hours a day. Infants in Group 2 could be touched only by someone wearing gloves. After 4 weeks of these differing treatment conditions, you could compare the head circumferences of the babies in these different conditions. In this study, there are three things creating variability in infant's head circumference. The fact that measuring an infant's head circumference is hard to do accurately contributes to the amount of _______________ in this study.

   a. treatment differences variability

   b. individual differences variability

   c. measurement error variability

7. The fact that some infants had 6 hours of skin-to-skin contact per day and others had no skin-to-skin contact contributes to the amount of _______________ in this study.

   a. treatment differences variability

   b. individual differences variability

   c. measurement error variability

8. The fact that infants naturally differ in head size contributes to the amount of _______________ in this study.

   a. treatment differences variability

   b. individual differences variability

   c. measurement error variability

9. If we measured each of the following variables for every person in this class, which variable would have the most *measurement error variability*?

   a. Students' report of their parents' annual income.

   b. Parents' annual income recorded from official tax forms.

10. If we measured each of the following variables for every person in this class, which variable would have the least *individual differences variability*?

    a. Number of siblings a person has.

    b. Number of fingers a person has.

11. Understanding variability is important because some variables simply have more variability than others. For example, when considering high school students, which of the following variables would have the largest standard deviation (i.e., the most variability)?

    a. Annual income of parents.

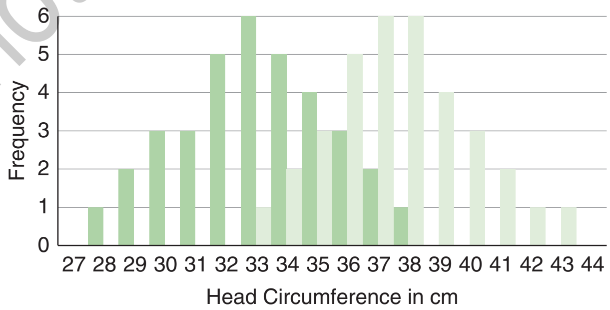
    b. Age.

12. Which of the following variables would have the smallest standard deviation (i.e., the least amount of variability) for high school students?

    a. Number of phone calls made in a day.

    b. Number of phones owned.

13. The figure below displays the head circumferences of 70 premature infants. Half of the infants were touched only by someone wearing gloves (*the darker bars*). The other half of the infants were touched only by someone who was not wearing gloves (*the lighter bars*).



In the previous figure, the variability created by the different treatments (i.e., touching infants while wearing gloves vs. touching without gloves) is depicted by the fact that

a. all the infants who were touched by someone wearing gloves do not have the same head circumference.

b. all the infants who were touched by someone not wearing gloves do not have the same head circumference.

c. the infants who were touched by someone not wearing gloves (*lighter bars*) tended to have larger head circumferences than infants who were touched by someone wearing gloves (*darker bars*).

14. In most research situations, there will be variability that is created by *measurement error*, *individual differences*, and *differing treatments*. In the study described above, the researcher expected that touch would result in faster growth. Thus, the researcher compares the mean head circumferences for the "skin-touched group" to the "gloved-touched group." Suppose that the mean head circumference for the skin-touched sample was 38 cm, and the mean for the other sample was 33 cm. You can't just look at these two numbers and conclude that direct skin touching facilitated infant growth because the variability between the sample means may have been created by (Select all that apply.)

    a. a treatment effect.

    b. individual differences.

    c. measurement error.

The answer to the above question is all of the above. You cannot just look at the mean from each group and conclude that the treatment created the difference in mean head sizes because it is possible that the two groups of infants began the experiment with different head sizes. This potential problem could be created by sampling error, specifically getting two samples of infants with different mean head sizes by chance. Later in this course you will learn how to calculate the probability that this sampling error problem occurred.

15. Which of the following is a correct conclusion for these data?

    a. It would never be possible to conclude with any confidence that the treatment probably created the difference in head size.

    b. By using statistical procedures you will learn later in the course, you could determine the probability that the difference in mean head size between the two groups was created by sampling error, and if the probability is small, then you could be confident that the treatment created the differences in head size.

Activity 2.3

Does Neonatal Massage Increase Infants' Weight?

1. You want to conduct a study to determine if neonatal massage increases infant growth by comparing the weights of infants before and after 1 month of massage therapy. Obviously, if the massage therapy is effective, you would expect infants to weigh more after the therapy. However, even in the absence of massage therapy, you would expect infants' weights to increase after 1 month. Why?

    a. Measurement error will decrease over the month.

    b. Measurement error will increase over the month.

    c. Infants grow over the month.

2. Clearly, you can't just measure weight before and after massage therapy. A better study would randomly assign infants to one of two groups where one group receives the massage therapy and another group does not (a control group). In this case, you would expect the mean weights of both groups to increase after 1 month but, if the massage therapy is effective, the massage group should gain _______________ weight than the control group.

    a. more

    b. less

3. Your study used a massage therapy group and a control group that received no massage therapy. The massage group gained more weight (in grams) $M = 589$ ($SD = 121$) than the control group $M = 501$ ($SD = 125$). Which of the following is the best scientific conclusion?

    a. These data prove that massage therapy works!

    b. These data suggest that massage therapy works but there might be other explanations for why the massage group gained more weight than the control group. It is best to be cautious and collect additional data, perhaps using more-rigorous control procedures that help rule out alternative explanations for the results.

4. Although having a no–massage therapy control group is a huge improvement over having no control group at all, your study's methodological rigor could be improved further. For example, parents and hospital workers might have treated the infants differently because they knew which infants were receiving massage therapy (e.g., more eye contact, more talking). These unintentional treatment differences might create different weight gain in the groups. A clever study

controlled for these possibilities with a placebo control (Ang et al., 2012). In this study, infants were randomly assigned to massage therapy or placebo therapy. One group received massages from a certified infant massage therapist. The therapist stood behind a screen so that the baby could not see the person giving the massage. For the placebo group, the massage therapist stood behind the screen but did not physically contact the infant. No parents or hospital workers knew which infants received massage therapy. The amount of all infants' weight gain was recorded. This example illustrates why you should consider the methodological rigor of the study generating the statistical results. The same mean difference might be interpreted very differently based on the methodology that generated it. Which of the following would provide the strongest evidence for massage therapy?

a. A mean weight gain for a sample of infants after getting massage therapy.
b. A larger weight gain for a sample of infants that got massage therapy compared to a sample that did not receive massage therapy.
c. A larger weight gain for a sample of infants that got massage therapy compared to a sample that got a placebo control treatment.
d. All of these are equally strong evidence in support of massage therapy.

5. The results of one study (Ang et al., 2012) indicated that infants in the massage group gained more weight than infants in the placebo control group. Suppose you replicated this study using the same massage and control procedures. In your study, you include several different measures of infant health and well-being, including the Apgar score. Health-care professionals use this scale to assess infants' health on five different factors (Appearance, Pulse, Grimace, Activity, and Respiration). For example, to assess Respiration, babies are rated as 2 = Normal rate and effort, good cry; 1 = slow or irregular breathing, weak cry; 0 = absent, no breathing. What scale of measurement is used for this Respiration variable (nominal, ordinal, or interval/ratio)?
6. Is this variable discrete or continuous?
7. To answer the following questions, you will need the data file "Activity 2.3 Infant Massage Data." Open this file in your statistical software. Babies are rated on the five different Apgar factors and then their scores are summed to yield a single measure with scores between 0 and 10, with higher numbers being more favorable. Use statistical software to create a bar graph for the Apgar scores in the ***control group***. Do the data appear to be approximately normally distributed, negatively skewed, or positively skewed?

8. Use statistical software to create a bar graph of Apgar scores for the ***massage therapy group***. Are the data in this group approximately normally distributed, positively skewed, or negatively skewed?

9. Do the distributions of Apgar scores (i.e., bar graphs) for the control and massage groups look pretty similar or very different?

10. Statistical software easily generates measures of central tendency (i.e., mean, median, and mode). The values for the ***control group*** are presented below. What is the best measure of central tendency for these ordinal data (Apgar scores)?

Descriptive Statistics

	Apgar Control
Valid	56
Missing	64
Mean	7.000
Median	7.000
Mode	8.000
Std. Deviation	2.018
Minimum	1.000
Maximum	10.000

11. Use statistical software to find the median for the Apgar scores in the ***massage therapy group***. Look at the graphs and medians for both groups. These data were collected before the treatment began, therefore the two groups should produce similar graphs and medians. Do the two groups have similar Apgar scores?

12. In your study, you also recorded the weights of the infants at birth (before treatment began) in kilograms. What is the scale of measurement for weight (nominal, ordinal, interval/ratio)?

13. Create a boxplot for the starting weights of babies in the ***control group***. Is the value marked by the center line in the box on the boxplot the mean, median, or mode?

14. In the boxplot of starting weights in the control group, the bottom line on the box represents the ________________ percentile and the top line on the box represents the ________________ percentile.

15. Use software to create a histogram and boxplot of starting weights for the ***massage therapy group*** before any treatments were administered. Based on the graphs, does the massage therapy group look similar to the control group before treatment?

16. The descriptive statistics for both groups are shown below. What measure of central tendency should you use for the starting weights of the infants? The mean, median, or mode?

Descriptive Statistics

	Starting Weight Control	Starting Weight Massage
N	58	56
Missing	62	64
Mean	3.03	3.10
Median	3.10	3.10
Minimum	2.00	2.10
Maximum	4.00	4.00

17. Based on the descriptive statistics, do the two groups appear to have similar starting weights?
    a. Yes
    b. No

18. Based on their Apgar scores and starting weights, do the two groups of infants appear to be about equal in health before they started treatment?
    a. Yes, they appear to be about equal in health.
    b. No, the massage therapy group appears to be in better health.
    c. No, the control group appears to be in better health.

19. After 1 month of treatment, you weigh all the infants again and record the amount of weight they gained (in kilograms). Use statistical software to generate a histogram, bar graph, and descriptive statistics for the amount of weight gained in the ***control group***. Based on this information, what measure of central tendency should you use to describe this sample?

20. What is the range of weight gain in the ***control group***?

21. What is the typical distance between individuals' weight gain and the mean weight gain in the ***control group***?

22. Use statistical software to obtain a boxplot, histogram, and descriptive statistics for the amount of weight gained in the ***massage therapy group***. Choose the best measure of central tendency for these data and record the value.

23. What is the standard deviation for the ***massage therapy group***?

24. Which of the following statements is the best interpretation of the standard deviation of the scores?

    a. The amount of sampling error in a study.

    b. The typical distance between scores and the mean.

    c. The typical distance between the sample mean and the population mean.

    d. The typical distance between all the scores in the data set.

25. Compare the graphs and the means for the massage therapy and control groups. Which group appears to have gained the most weight?

    a. The massage group.

    b. The control group.

26. Explain why a higher mean weight gain for the massage group does not ***prove*** that massage therapy is effective.

    a. The massage group might have a higher weight because of sampling error.

    b. The massage group might have a higher weight because of a methodological flaw with the study.

    c. The massage group might have a higher weight because of both sampling error and a methodological flaw with the study.

27. Does it look like massage might be effective? How compelling is the evidence? Construct a scientific conclusion based on the evidence. Be sure to explicitly mention the evidence on which you are basing your conclusion.

Activity 2.4

Choosing Measure of Central Tendency and Variability

For each of the following situations determine if you should use the standard deviation or range as the measure of variability.

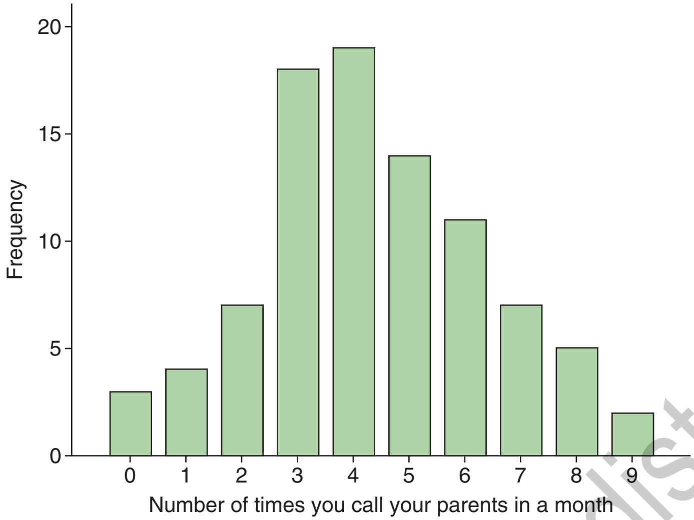
1. The variable is ordinal.
2. The variable is interval/ratio.
3. The median is reported as the measure of central tendency.

For each of the following situations, determine if you should use the mean, median, or mode as the measure of central tendency. The variable is

4. nominal.
5. ordinal.
6. interval/ratio.
7. interval/ratio but is highly skewed.
8. interval/ratio but there are outliers.
9. height of children in fifth grade; the heights are approximately normally distributed with no outliers.
10. scores on an exam scored as A, B, C, D, or F.
11. scores on an exam scored as a percentage correct (0–100%); most people earned scores between 100 and 78, but two people earned scores below 20.

For the following problems, determine which measure of central tendency is the most appropriate. Be sure to consider the scale of measurement as well as the shape of the distribution.

12. These data came from a sample of college students who attend college at least 100 miles away from their parents. They were asked how many times they called their parents in a typical month. The mean, median, and mode were 4.34, 4, and 4, respectively. Which measure of central tendency should be used and why?
    a. The mean because the data are interval/ratio and the distribution is not very skewed.
    b. The mode because the data are discrete.
    c. The median because the data are ordinal.
    d. All measures of central tendency are equally appropriate for these data.

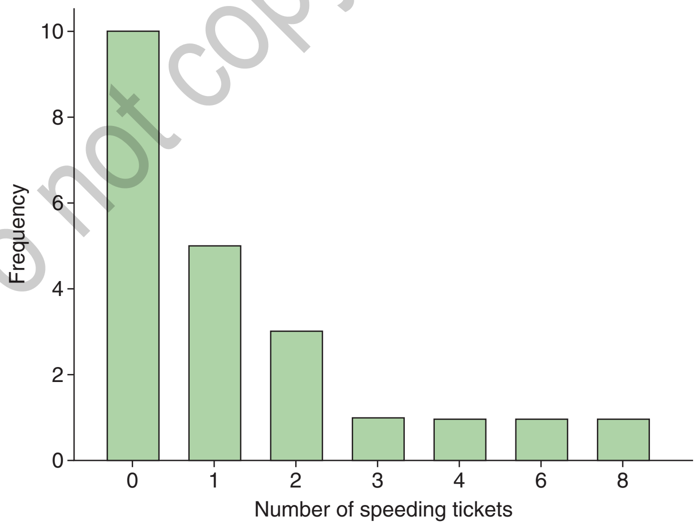


13. These data came from people attending a stock car race in Indiana. They were asked how many speeding tickets they had received in the previous 2 years. The mean, median, and mode were 1.45, 1, and 0, respectively. Which measure of central tendency should be used and why? See graph below.

a. The median because the data are ordinal.

b. The mean because the data are interval/ratio.

c. The median because the data are interval/ratio, but skewed.

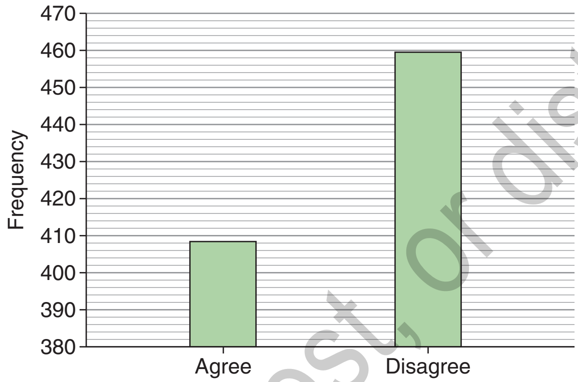


14. These data came from clinical psychologists who were attending a professional conference. They responded to the question, "Do you agree that patients' memories of past events are improved by hypnosis?" What is the best measure of central tendency for these data and why? See graph below.

   a. The mode because the data are ordinal.

   b. The median because the data are skewed.

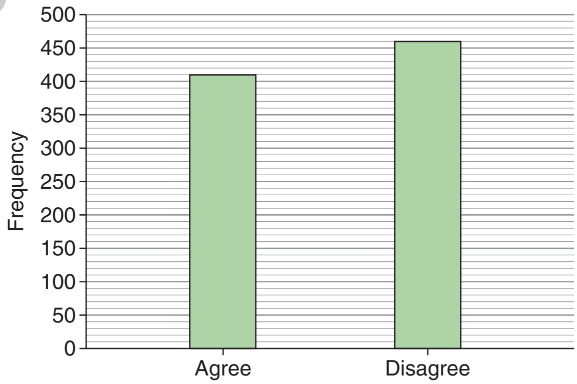
   c. The mode because the data are nominal.



15. This graph presents the same data that were presented in the above graph. Why do the graphs look so different? See graph below.

   a. They look different because the range of values on the y-axis is different in the two graphs.

   b. They look different because the second graph uses an interval/ratio scale while the first uses a nominal scale.



16. Which do you think is a more accurate representation of the data a nd why? (It is important to note that memories are *not* improved by hypnosis.)

    a. The first graph is better because it highlights the difference between the agree and disagree responses.

    b. The second graph is better because it shows the true range of values.

17. These data were obtained from a sample of people in the supermarket. They were asked how many miles they ran in the previous week. The mean, median, and mode were 3.37, 5, and 0, respectively. In this case, no one number accurately describes the distribution of scores. In fact, one could argue that presenting only a measure of central tendency would be misleading. Which of the following is the best summary of these data? See graph below.

    a. The mean number of miles respondents reported running was 3.37 with a median of 5 and a mode of 0.

    b. Many respondents reported that they did not run that week. Of those who did run, the reported miles ranged between 5 and 9, with a mean of 6.57.

    c. The number of miles people reported running varied between 0 and 9, with a mean of 3.37 and a median of 5.