The Mean, Variance, Standard Deviation, and *Z* Scores

Chapter Outline

The purpose of descriptive statistics is to make a group of scores understandable. In this chapter, we consider the main statistical techniques for describing a group of scores with numbers. First, you can describe a group of scores in terms of a *representative* (or *typical*) *value*, such as an average. A representative value gives the *central tendency* of a group of scores. A representative value is a simple way, with a single number, to describe a group of scores (and there may be hundreds or even thousands of scores). The main representative value we focus on is the *mean*. Next, we focus on ways of describing how spread out the numbers are in a group of scores. In other words, we consider the amount of variation, or *variability*, among the scores. The two measures of variability you will learn about are called the *variance* and *standard deviation*. Finally, we show you how to describe a particular score in terms of how much that score varies from the average. To do this, you will learn how to combine the mean and standard deviation to create a *Z score*.

TIP FOR SUCCESS

Before beginning this chapter, you should be sure that you are comfortable with the key terms of variable, score, and value.

In this chapter, you will use statistical formulas. Such formulas are not here to confuse you. Hopefully, you will come to see that they actually simplify things and provide a very straightforward, concise way of describing statistical procedures. To help you grasp what such formulas mean in words, whenever we present formulas, we always also give the "translation" in ordinary English.

Representative Values

The representative value of a group of scores (a distribution) refers to the middle of the group of scores. You will learn about three representative values: *mean*, *mode*, and *median*. Each uses its own method to come up with a single number describing the middle of a group of scores. We start with the mean, the most commonly used measure of the representative value of a group of scores.

The Mean

Usually, the best measure of the representative value of a group of scores is the ordinary average, the sum of all the scores divided by the number of scores. In statistics, this is called the **mean.** Suppose that a political scientist does a study on years of experience in elected office. As part of this research, the political scientist finds out the number of years served by mayors of the 10 largest cities in a particular region. The numbers of years served were as follows:

7, 8, 8, 7, 3, 1, 6, 9, 3, 8

The mean of these 10 scores is 6 (the sum of 60 years served divided by 10 mayors). That is, on average, these 10 mayors had served 6 years in office. The information for the 10 mayors is thus summarized by this single number, 6.

You can think of the mean as a kind of balancing point for the distribution of scores. Try it by visualizing a board balanced over a log, like a rudimentary seesaw. Imagine piles of blocks set along the board according to their values, one for each score in the distribution. (This is a little like a histogram made of blocks.) The mean is the point on the board where the weight of the blocks on one side balances exactly with the weight on the other side. Figure 1 shows this for our 10 mayors.

Mathematically, you can think of the mean as the point at which the total distance to all the scores above that point equals the total distance to all the scores below that point. Let's first figure the total distance from the mean to all of the scores above the mean for the mayors' example shown in Figure 1. There are two scores of 7, each of which is 1 unit above 6 (the mean). There are three scores of 8, each of which is 2 units above 6. And, there is one score of 9, which is 3 units above 6. This gives a total

mean (*M*) Arithmetic average of a group of scores; sum of the scores divided by the number of scores.

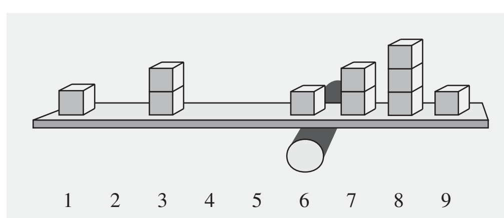


Figure 1 Mean of the distribution of the numbers of years in office for 10 mayors, illustrated using blocks on a board balanced on a log.

distance of 11 units $(1 + 1 + 2 + 2 + 2 + 3)$ from the mean to all the scores above the mean. Now, let's look at the scores below the mean. There are two scores of 3, each of which is 3 units below 6 (the mean). And there is one score of 1, which is 5 units below 6. This gives a total distance of 11 units $(3 + 3 + 5)$ from the mean to all of the scores below the mean. Thus, you can see that the total distance from the mean to the scores above the mean is the same as the total distance from the mean to the scores below the mean. The scores above the mean balance out the scores below the mean (and vice versa).

Some other examples are shown in Figure 2. Notice that there doesn't have to be a block right at the balance point. That is, the mean doesn't have to be a score actually in the distribution. The mean is the average of the scores, the balance point. The mean can be a decimal number, even if all the scores in the distribution have to be whole numbers (a mean of 2.3 children, for example). For each distribution in Figure 2, the total distance from the mean to the scores above the mean is the same as the total distance from the mean to the scores below the mean. (By the way, this analogy to blocks on a board, in reality, would work out precisely only if the board had no weight of its own.)

TIP FOR SUCCESS

Think of each formula as a statistical recipe, with statistical symbols as ingredients. Before you use each formula, be sure you know what each symbol stands for. Then, carefully follow the formula to come up with the end result.

Formula for the Mean and Statistical Symbols The rule for figuring the mean is to add up all the scores and divide by the number of scores. Here is how this can be written as a formula:

$$M = \frac{\sum X}{N} \quad (1)$$

The mean is the sum of the scores divided by the number of scores.

M is a symbol for the mean. An alternative symbol, $\overline{X}$ ("*X*-bar"), is sometimes used. However, *M* is most commonly used in published research articles. In fact, you should know that there is not a general agreement for many of the symbols used in statistics. (In this text, we generally use the symbols most widely found in research publications.)

Σ, the capital Greek letter "sigma," is the symbol for "sum of." It means "add up all the numbers" for whatever follows. It is the most common special arithmetic symbol used in statistics.

M Mean.

Σ Sum of; add up all the scores following.

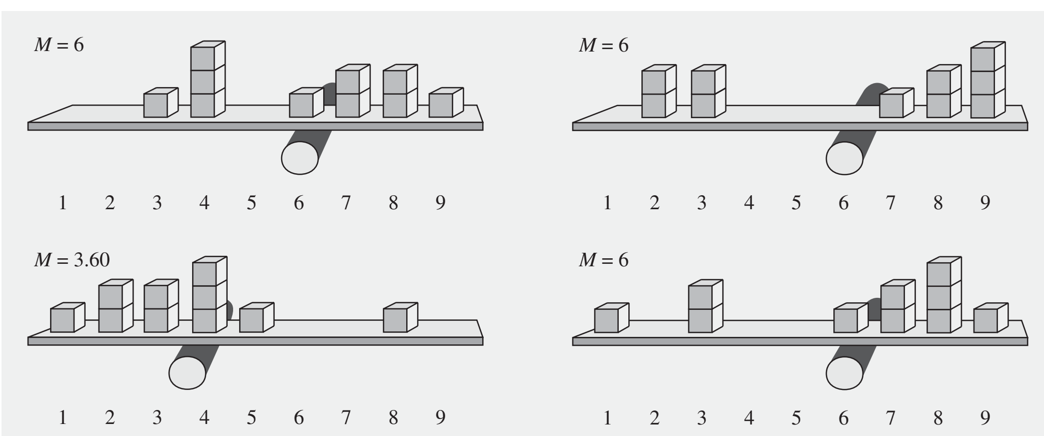


Figure 2 Means of various distributions illustrated with blocks on a board balanced on a log.

X stands for the scores in the distribution of the variable *X*. We could have picked any letter. However, if there is only one variable, it is usually called *X*. We use formulas with more than one variable. In those formulas, we use a second letter along with *X* (usually *Y*) or subscripts (such as X_1 and X_2).

$\sum X$ is "the sum of *X*." That is, this tells you to add up all the scores in the distribution of the variable *X*. Suppose *X* is the number of years in office in our example of 10 mayors: $\sum X$ is $7 + 8 + 8 + 7 + 3 + 1 + 6 + 9 + 3 + 8$, which is 60.

N stands for number—the number of scores in a distribution. In our example, there are 10 scores. Thus, *N* equals 10.

Overall, the formula says to divide the sum of all the scores for the variable *X* by the total number of scores, *N*. In our example, this tells us that we divide 60 by 10. Put in terms of the formula,

$$M = \frac{\sum X}{N} = \frac{60}{10} = 6$$

Additional Examples of Figuring the Mean. The stress ratings of the 30 students in the first week of their statistics class (based on Aron et al., 1995) are:

8, 7, 4, 10, 8, 6, 8, 9, 9, 7, 3, 7, 6, 5, 0, 9, 10, 7, 7, 3, 6, 7, 5, 2, 1, 6, 7, 10, 8, 8

We summarized all these scores into a frequency table (Table 1–3). You can now summarize all this information as a single number by figuring the mean. Figure the mean by adding up all the stress ratings and dividing by the number of stress ratings. That is, you add up the 30 stress ratings: $8 + 7 + 4 + 10 + 8 + 6 + 8 + 9 + 9 + 7 + 3 + 7 + 6 + 5 + 0 + 9 + 10 + 7 + 7 + 3 + 6 + 7 + 5 + 2 + 1 + 6 + 7 + 10 + 8 + 8$, for a total of 193. Then you divide this total by the number of scores, 30. In terms of the formula,

$$M = \frac{\sum X}{N} = \frac{193}{30} = 6.43$$

TIP FOR SUCCESS

When an answer is not a whole number, we suggest that you use two more decimal places in the answer than for the original numbers. In this example, the original numbers did not use decimals, so we rounded the answer to two decimal places.

This tells you that the average rating was 6.43 (after rounding off). This is clearly higher than the middle of the 0–10 scale. You can also see this on a graph. Think again of the histogram as a pile of blocks on a board and the mean of 6.43 as the point where the board balances on a fulcrum (see Figure 3). This single representative value simplifies the information in the 30 stress scores.

Similarly, consider the example of students' social interactions (McLaughlin-Volpe, Aron, & Reis, 2001). The actual number of interactions over a week for the 94 students is not listed here. We can take those 94 scores, add them up, and divide by 94 to figure the mean:

$$M = \frac{\sum X}{N} = \frac{1635}{94} = 17.39$$

This tells us that during this week these students had an average of 17.39 social interactions. Figure 4 shows the mean of 17.39 as the balance point for the 94 social interaction scores.

Steps for Figuring the Mean. Figure the mean in two steps.

❶ **Add up all the scores.** That is, figure $\sum X$.
❷ **Divide this sum by the number of scores.** That is, divide $\sum X$ by *N*.

X Scores in the distribution of the variable *X*.

N Number of scores in a distribution.

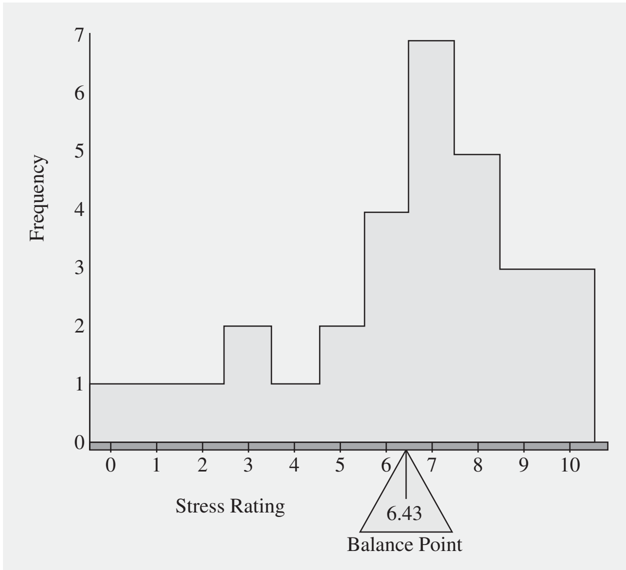


Figure 3 Analogy of blocks on a board balanced on a fulcrum showing the mean for 30 statistics students' ratings of their stress level. (Data based on Aron, Paris, and Aron, 1995.)

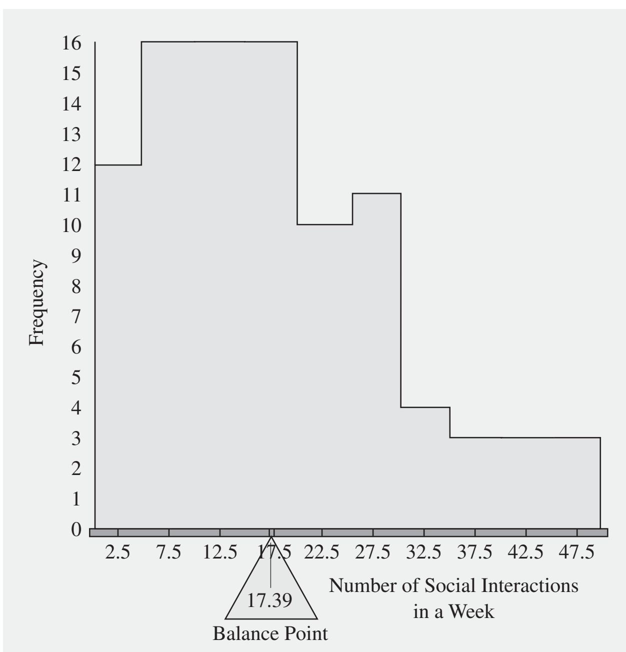


Figure 4 Analogy of blocks on a board balanced on a fulcrum showing the mean for number of social interactions during a week for 94 college students. (Data from McLaughlin-Volpe et al., 2001.)

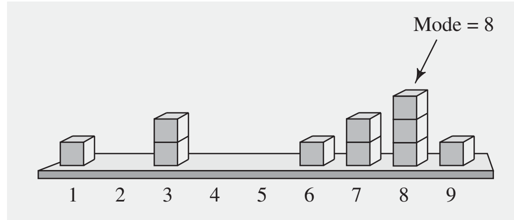


Figure 5 The mode as the high point in a distribution's histogram, using the example of the number of years in office served by 10 mayors.

The Mode

The **mode** is another measure of the representative (or typical) value in a group of scores. The mode is the most common single value in a distribution. In our mayors' example, the mode is 8. This is because there are three mayors with 8 years served in office and no other number of years served in office with as many mayors. Another way to think of the mode is that it is the value with the largest frequency in a frequency table, the high point or peak of a distribution's histogram (as shown in Figure 5).

In a perfectly symmetrical unimodal distribution, the mode is the same as the mean. However, what happens when the mode and the mean are not the same? In that situation the mode is usually not a very good representative value for scores in the distribution. In fact, sometimes researchers compare the mode to the mean to show that the distribution is *not* perfectly symmetrical. Also, the mode can be a particularly poor representative value because it does not reflect many aspects of the distribution. For example, you can change some of the scores in a distribution without affecting the mode—but this is not true of the mean, which is affected by any change in the distribution (see Figure 6).

On the other hand, the mode *is* the usual way of describing the representative value for a nominal variable. For example, if you know the religions of a particular group of people, the mode tells you which religion is the most frequent. However, when it comes to the numerical variables that are most common in behavioral and social science research, the mode is rarely used.

The Median

TIP FOR SUCCESS

When figuring the median, remember the first step is to line up the scores from lowest to highest. Forgetting to do this is the most common mistake students make when figuring the median.

Another different measure of the representative value of a group of scores is the **median.** If you line up all the scores from lowest to highest, the middle score is the median. As shown in Figure 7, if you line up the numbers of years in office from lowest to highest, the fifth and sixth scores (the two middle ones) are both 7s. Either way, the median is 7 years.

When you have an even number of scores, the median is between two scores. In that situation, the median is the average (the mean) of the two scores.

Steps for Finding the Median. Finding the median takes three steps.

❶ **Line up all the scores from lowest to highest.**

❷ **Figure how many scores there are to the middle score, by adding 1 to the number of scores and dividing by 2.** For example, with 29 scores, adding 1 and dividing by 2 gives you 15. The 15th score is the middle score. If there are 50 scores, adding 1 and dividing by 2 gives you 25.5. Because there are no half scores, the 25th and 26th scores (the scores either side of 25.5) are the middle scores.

mode Value with the greatest frequency in a distribution.

median Middle score when all the scores in a distribution are arranged from lowest to highest.

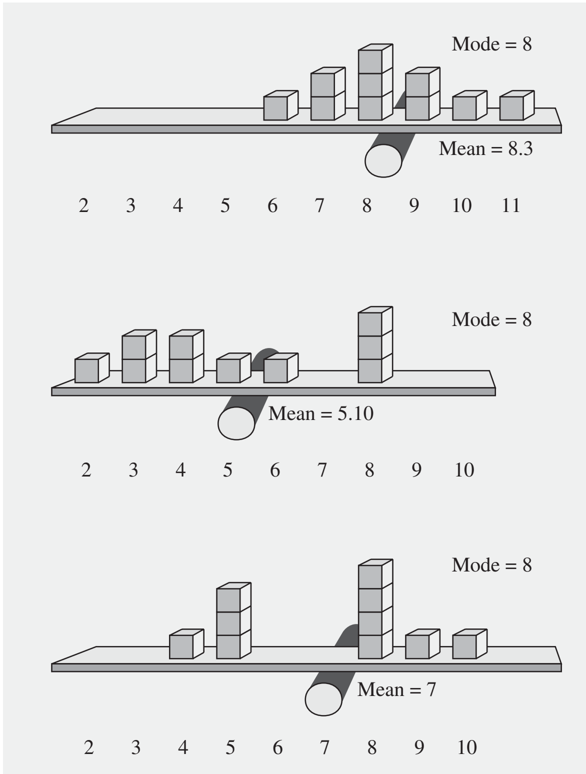


Figure 6 The effect on the mean and on the mode of changing some scores, using the example of the number of years in office served by 10 mayors.

❸ **Count up to the middle score or scores.** If you have one middle score, this is the median. If you have two middle scores, the median is the average (the mean) of these two scores.

Comparing Representative Values

Sometimes, the median is better than the mean (or mode) as a representative value for a group of scores. This happens when a few extreme scores would strongly affect the mean but would not affect the median. For example, suppose that among the 100 families on a banana plantation in Central America, 99 families have an annual income of \$500 and one family (the owner's) has an annual income of \$450,500. The mean family income on this plantation would be \$5,000 ($99 \times 500 = 49{,}500$; $49{,}500 + 450{,}500 = 500{,}000$; $500{,}000/100 = 5{,}000$). No family has an income

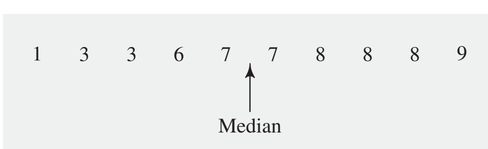


Figure 7 The median is the middle score when scores are lined up from lowest to highest, using the example of the number of years in office served by 10 mayors.

even close to \$5,000, so this number is completely misleading. The median income in this example would be \$500—an income much more representative of whomever you would meet if you walked up to someone randomly on the plantation.

As this example illustrates, behavioral and social scientists use the median as a descriptive statistic mainly in situations where there are a few extreme scores that would make the mean unrepresentative of most of the scores. An extreme score like this is called an **outlier.** (In this example, the outlier was much higher than the other scores, but in other cases an outlier may be much lower than the other scores in the distribution.) There are also times when the median is used as part of more complex statistical methods. However, unless there are extreme scores, behavioral and social scientists almost always use the mean as the measure of the representative value of a group of scores. In fact, as you will learn, the mean is a fundamental building block for most other statistical techniques.

How are you doing?

1. Name and define three measures of the representative value of a group of scores.
2. Write the formula for the mean and define each of the symbols.
3. Figure the mean of the following scores: 2, 3, 3, 6, and 6.
4. For the following scores, find (a) the mean; (b) the mode; (c) the median: 5, 3, 2, 13, 2. (d) Why is the mean different from the median?

Answers

1. The mean is the ordinary average, the sum of the scores divided by the number of scores. The mode is the most frequent score in a distribution. The median is the middle score, that is, if you line the scores up from lowest to highest, it is the halfway score.
2. $M = (\sum X)/N$. M is the mean; $\sum$ is the symbol for "sum of"—add up all the scores that follow; X stands for the variable whose scores you are adding up; N is the number of scores.
3. $M = (\sum X)/N = (2 + 3 + 3 + 6 + 6)/5 = 20/5 = 4$.
4. (a) The mean is 5; (b) the mode is 2; (c) the median is 3. (d) The mean is different from the median because the extreme score (13) makes the mean much higher than the median.

Variability

Researchers also want to know how spread out the scores are in a distribution. This shows the amount of variability in the distribution. For example, suppose you were asked, "How old are the students in your statistics class?" At a city-based university with many returning and part-time students, the mean age might be 34 years. You could answer, "The average age of the students in my class is 34." However, this would not tell the whole story. You could, for example, have a mean of 34 because every student in the class was exactly 34 years old. If this is the case, the scores in the distribution are not spread out at all. In other words, there is no variation, or *variability*, among the scores. You could also have a mean of 34 because exactly half the class members were 24 and the other half 44. In this situation, the distribution is much more spread out. There is considerable variability among the scores.

You can think of the variability of a distribution as the amount of spread of the scores around the mean. Distributions with the same mean can have very different

outlier Score with an extreme value (very high or very low) in relation to the other scores in the distribution.

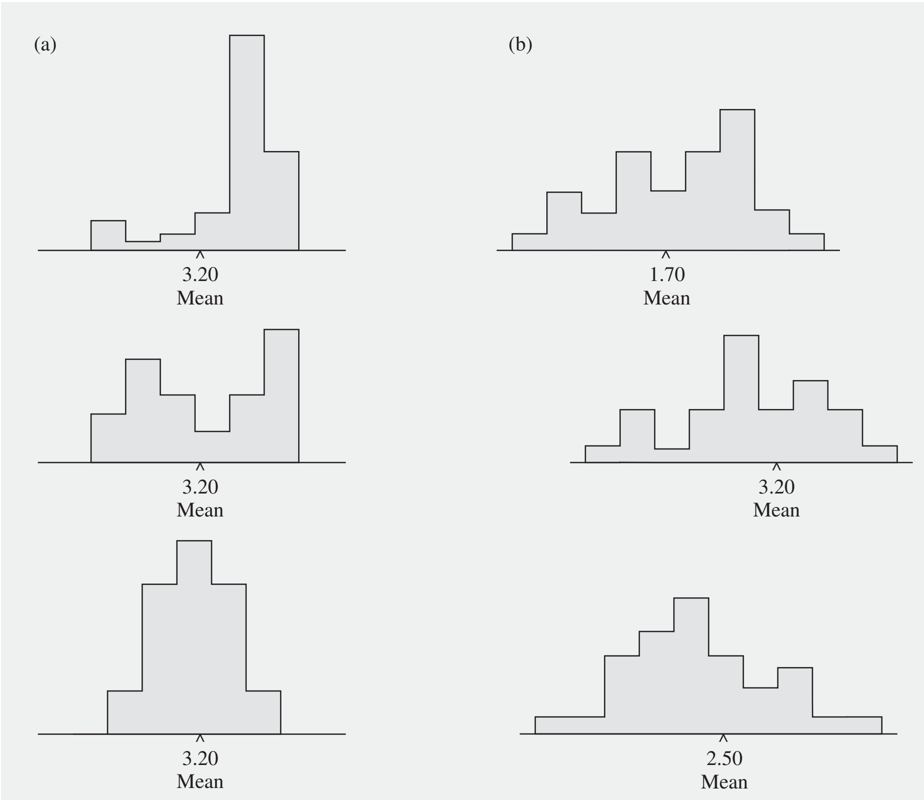


Figure 8 Examples of distributions with (a) the same mean but different amounts of spread, and (b) different means but the same amount of spread.

amounts of spread around the mean; Figure 8a shows histograms for three different frequency distributions with the same mean but different amounts of spread around the mean. Also, distributions with different means can have the same amount of spread around the mean. Figure 8b shows three different frequency distributions with different means but the same amount of spread. So, although the mean provides a representative value of a group of scores, it doesn't tell you about the variability (or spread) of the scores. You will now learn about two measures of the variability of a group of scores: the *variance* and *standard deviation*.

The Variance

The **variance** of a group of scores tells you how spread out the scores are around the mean.[1] To be precise, the variance is the average of each score's squared difference from the mean.

variance Measure of how spread out a set of scores are; average of the squared deviations from the mean.

[1]This section focuses on the variance and standard deviation as measures of variation (or spread). There is also another way to describe the variation of a group of scores, the *range*—the highest score minus the lowest score. Suppose that for a particular class of students the highest score on a midterm is 98 and the lowest score on the midterm is 60; the range is 38 (that is, $98 - 60 = 38$). Researchers rarely use the range because it is a very crude way of describing the variation. It is crude because it considers only two scores from the group of scores and does not take into account how clumped together or spread out the scores are within the range.

Here are the four steps to figure the variance.

❶ **Subtract the mean from each score.** This gives each score's **deviation score,** which is how far away the score is from the mean.
❷ **Square each of these deviation scores** (multiply each by itself). This gives each score's **squared deviation score.**
❸ **Add up the squared deviation scores.** This total is called the **sum of squared deviations.**
❹ **Divide the sum of squared deviations by the number of scores.** This gives the average (the mean) of the squared deviations, called the variance.

This procedure may seem a bit awkward or hard to remember at first, but it works quite well. Suppose one distribution is more spread out than another. The more spread-out distribution has a larger variance because being spread out makes the deviation scores bigger. If the deviation scores are bigger, the squared deviation scores also are bigger. Thus, the average of the squared deviation scores (the variance) is bigger.

In the example of the class in which everyone was exactly 34 years old, the variance would be exactly 0. That is, there would be no variance (which makes sense, as there is no variability among the ages). (In terms of the numbers, each person's deviation score would be $34 - 34 = 0$; 0 squared is 0. The average of a bunch of zeros is 0.) By contrast, the class of half 24-year-olds and half 44-year-olds would have a rather large variance of 100. (The 24-year-olds would each have deviation scores of $24 - 34 = -10$. The 44-year-olds would have deviation scores of $44 - 34 = 10$. All the squared deviation scores, which are -10 squared or 10 squared, would come out to 100. The average of all 100s is 100.)

The variance is extremely important in many statistical procedures you will learn about later. However, the variance is rarely used as a *descriptive statistic*. This is because the variance is based on *squared* deviation scores, which do not give a very easy-to-understand sense of how spread out the actual, nonsquared scores are. For example, it is clear that a class with a variance of 100 has a more spread-out distribution than one whose variance is 10. However, the number 100 does not give an obvious insight into the actual variation among the ages, none of which is anywhere near 100.

deviation score Score minus the mean.

squared deviation score Square of the difference between a score and the mean.

sum of squared deviations Total over all the scores of each score's squared difference from the mean.

standard deviation Square root of the average of the squared deviations from the mean; the most common descriptive statistic for variation; approximately the average amount that scores in a distribution vary from the mean.

The Standard Deviation

The most widely used way of *describing* the spread of a group of scores is the **standard deviation.** The standard deviation is directly related to the variance and is figured by taking the square root of the variance. There are two steps in figuring the standard deviation.

❶ **Figure the variance.**
❷ **Take the square root.** The standard deviation is the *positive* square root of the variance. (Any number has both a positive and a negative square root. For example, the square root of 9 is both +3 and −3.)

If the variance of a distribution is 100, the standard deviation is 10. If the variance is 16, the standard deviation is 4.

The variance is about squared deviations from the mean. Therefore, its square root, the standard deviation, is about direct, ordinary, nonsquared deviations from the mean. *Roughly speaking, the standard deviation is the average amount that scores differ from the mean.* For example, consider a class where the ages have a standard deviation of 10 years. This tells you that the ages are spread out, on average, about 10 years in each direction from the mean. Knowing the standard deviation gives you a general sense of the degree of spread.

TIP FOR SUCCESS

A common error when figuring the standard deviation is to jump straight from the sum of squared deviations to the standard deviation (by taking the square root of the sum of squared deviations). Remember, before finding the standard deviation, you first have to figure the variance (by dividing the sum of squared deviations by the number of scores, *N*). Then take the square root of the variance to find the standard deviation.

The standard deviation usually is not *exactly* the average amount that scores differ from the mean. To be precise, the standard deviation is the square root of the average of the scores' squared deviations from the mean. This squaring, averaging, and then taking the square root usually gives a slightly different result from simply averaging the scores' deviations from the mean. Still, the result of this approach has technical advantages to outweigh this slight disadvantage of giving only an approximate description of the average variation from the mean.

Formulas for the Variance and the Standard Deviation. We have seen that the variance is the average squared deviation from the mean. Here is the formula for the variance:

$$SD^2 = \frac{\Sigma(X - M)^2}{N} \qquad (2)$$

The variance is the sum of the squared deviations of the scores from the mean, divided by the number of scores.

SD*²** is the symbol for the variance. ***SD is short for *standard deviation*. The symbol SD^2 emphasizes that the variance is the standard deviation squared.

The top part of the formula is the *sum of squared deviations*. *X* is for each score and *M* is the mean. Thus, $X - M$ is the score minus the mean, the deviation score. The ^{2} tells you to square each deviation score. Finally, the sum sign (Σ) tells you to add up all these squared deviation scores. The bottom part of the formula tells you to divide the sum of squared deviation scores by *N*, the number of scores.

The standard deviation is the square root of the variance. So, if you already know the variance, the formula is

$$SD = \sqrt{SD^2} \qquad (3)$$

The standard deviation is the square root of the variance.

SD^2 Variance.

SD Standard deviation.

Examples of Figuring the Variance and the Standard Deviation. Table 1 shows the figuring for the variance and standard deviation for our mayors' example.

Table 1 Figuring of Variance and Standard Deviation in the Example of Number of Years Served by 10 Mayors

Score (Number of Years Served)	−	Mean score (Mean Number of Years Served)	=	Deviation score	Squared Deviation score
7		6		1	1
8		6		2	4
8		6		2	4
7		6		1	1
3		6		−3	9
1		6		−5	25
6		6		0	0
9		6		3	9
3		6		−3	9
8		6		2	4
				Σ: 0	66

Variance $= SD^2 = \dfrac{\Sigma(X - M)^2}{N} = \dfrac{66}{10} = 6.60$

Standard deviation $= SD = \sqrt{SD^2} = \sqrt{6.60} = 2.57$

TIP FOR SUCCESS

Notice in Table 1 that the deviation scores (shown in the third column) add up to 0. The sum of the deviation scores is *always* 0 (or very close to 0, allowing for rounding error). This is because, as you learned earlier in the chapter, the mean is the balancing point of a distribution (where the total distance from the mean to the scores above the mean is the same as the total distance from the mean to the scores below the mean). So, to check your figuring, always sum the deviation scores. If they do not add up to 0, do your figuring again!

TIP FOR SUCCESS

Always check that your answers make *intuitive sense*. For example, looking at the scores for the mayors' example, a standard deviation—which, roughly speaking, represents the average amount the scores vary from the mean—of 2.57 makes sense. If your answer had been 421.23, however, it would mean that, on average, the number of years in office varied by more than 420 years from the mean of 6. Looking at the scores, that just couldn't be true.

(The table assumes we already have figured out the mean to be 6 years in office.) Usually, it is easiest to do your figuring using a calculator, especially one with a square root key. The standard deviation of 2.57 tells you that roughly speaking, on average, the number of years the mayors were in office varied by about 2½ from the mean of 6.

Table 2 shows the figuring for the variance and standard deviation for the example of students' numbers of social interactions during a week (McLaughlin-Volpe et al., 2001). To save space, the table shows only the first few and last few scores. Roughly speaking, this result indicates that a student's number of social interactions in a week varies from the mean (of 17.39) by an average of 11.49 interactions. This can also be shown on a histogram (see Figure 9).

Measures of variability, such as the variance and standard deviation, are heavily influenced by the presence of one or more outliers (extreme values) in a distribution. The scores in the mayors' example were 7, 8, 8, 7, 3, 1, 6, 9, 3, 8, and we figured the standard deviation of the scores to be 2.57. Now imagine that one additional mayor is added to the study who has been in office for 21 years. The standard deviation of the scores would now be 4.96, which is almost double the size of the standard deviation without this additional single score.

computational formula Equation mathematically equivalent to the definitional formula. It is easier to use for figuring by hand, but does not directly show the meaning of the procedure.

Computational and Definitional Formulas. In actual research situations, behavioral and social scientists must often figure the variance and the standard deviation for distributions with many scores, often involving decimals or large numbers. This can make the whole process quite time-consuming, even with a calculator. To deal with this problem, in the old days researchers developed various shortcuts to simplify the figuring. A shortcut formula of this type is called a **computational formula.**

TIP FOR SUCCESS

When figuring the variance and standard deviation, lay your working out as shown in Tables 1 and 2. This will help to ensure that you follow all of the steps and end up with the correct answers.

Table 2 Figuring the Variance and Standard Deviation for Number of Social Interactions during a Week for 94 College Students

Number of Interactions	−	Mean Number of Interactions	=	Deviation Score	Squared Deviation Score
48		17.39		30.61	936.97
15		17.39		−2.39	5.71
33		17.39		15.61	243.67
3		17.39		−14.39	207.07
21		17.39		3.61	13.03
.		.		.	.
.		.		.	.
.		.		.	.
35		17.39		17.61	310.11
9		17.39		−8.39	70.39
30		17.39		12.61	159.01
8		17.39		−9.39	88.17
26		17.39		8.61	74.13
				Σ: 0.00	12,406.44

$$\text{Variance} = SD^2 = \frac{\Sigma(X - M)^2}{N} = \frac{12{,}406.44}{94} = 131.98$$

$$\text{Standard deviation} = \sqrt{SD^2} = \sqrt{131.98} = 11.49$$

Source: Data from McLaughlin-Volpe et al. (2001).

INTERVAL	FREQUENCY
0 – 4	12
5 – 9	16
10 – 14	16
15 – 19	16
20 – 24	10
25 – 29	11
30 – 34	4
35 – 39	3
40 – 44	3
45 – 49	3

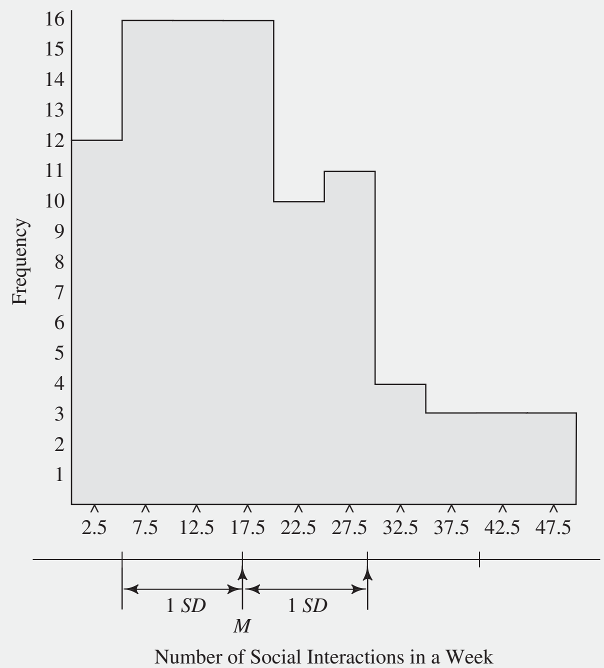


Figure 9 The standard deviation as the distance along the base of a histogram, using the example of number of social interactions in a week. (Data from McLaughlin-Volpe et al., 2001.)

The traditional computational formula for the variance of the kind we are discussing in this chapter is

$$SD^2 = \frac{\sum X^2 - [(\sum X)^2/N]}{N} \tag{4}$$

The variance is the sum of the squared scores minus the result of taking the sum of all the scores, squaring this sum and dividing by the number of scores, then taking this whole difference and dividing it by the number of scores.

$\sum X^2$ means that you square each score and then take the sum of these squared scores. However, $(\sum X)^2$ means that you add up all the scores first and then take the square of this sum. This formula is easier to use if you are figuring the variance for a lot of numbers by hand because you do not have to first find the deviation score for each score.

However, these days computational formulas are mainly of historical interest. They are used by researchers on rare occasions when computers with statistical software are not readily available to do the figuring. In fact, today, even many calculators are set up so that you need only enter the scores and press a button or two to get the variance and the standard deviation.

We give a few computational formulas so that you will have them if you someday do a research project with a lot of numbers and you don't have access to statistical software. However, we recommend *not* using the computational formulas when you are learning statistics. The problem is that the computational formulas tend to make it harder to understand the *meaning* of what you are figuring. It is much better to use the regular formulas we give when doing the practice problems. These are the formulas designed to help you strengthen your understanding of what the figuring *means*. These usual formulas we give are called **definitional formulas.**

definitional formula Equation for a statistical procedure directly showing the meaning of the procedure.

Our purpose is to help you understand statistical procedures, not to turn you into a computer by having you memorize computational formulas you will rarely, if ever, use. To simplify the actual figuring, however, our practice problems generally use small groups of whole numbers.

The Importance of Variability in Behavioral and Social Sciences Research

Variability is an important topic in behavioral and social sciences research because much of that research focuses on explaining variability. We will use an example to show what we mean by "explaining variability." As you might imagine, different students experience different levels of stress with regard to learning statistics: Some experience little stress; for other students, learning statistics can be a source of great stress. So, in this example, explaining variability means identifying the factors that explain why students differ in the amount of stress they experience. Perhaps how much experience students have had with math explains some of the variability in stress. That is, according to this explanation, the differences (the variability) among students in amount of stress are partially due to the differences (the variability) among students in the amount of experience they have had with math. Thus, the variation in math experience partially explains, or accounts for, the variation in stress. Perhaps the degree to which students generally experience stress in their lives also partly explains differences among students' stress with regard to learning statistics.

The Variance as the Sum of Squared Deviations Divided by $N - 1$

Researchers often use a slightly different kind of variance. We have defined the variance as the average of the squared deviation scores. Using that definition, you divide the sum of squared deviation scores by the number of scores. But for many purposes it is better to define the variance as the sum of squared deviation scores *divided by 1 less than the number of scores*. That is, for those purposes, the variance is the sum of squared deviations divided by $N - 1$. (You use this $N - 1$ approach when you have scores from a particular group of people and you want to estimate what the variance would be for the larger group of people whom these individuals represent.)

The variances and standard deviations given in research articles are usually figured using the $N - 1$ approach. Also, when calculators or computers give the variance or the standard deviation automatically, they are usually figured in this way (for example, see the "Using SPSS" section at the end of this chapter). But don't worry. The approach you are learning in this chapter of dividing by N is entirely correct for our purpose here, which is to use descriptive statistics to describe the variation in a particular group of scores. It is also entirely correct for the material covered in the rest of this chapter (Z scores of the kind we are using). We mention this $N - 1$ approach now only so you will not be confused if you read about variance or standard deviation in other places or if your calculator or a computer program gives a surprising result. To keep things simple, we will not discuss the $N - 1$ approach in this chapter.

How are you doing?

1. (a) Define the variance. (b) Describe what it tells you about a distribution and how this is different from what the mean tells you.
2. (a) Define the standard deviation, (b) describe its relation to the variance, and (c) explain what it tells you approximately about a group of scores.
3. Give the full formula for the variance and indicate what each of the symbols mean.
4. Figure the (a) variance and (b) standard deviation for the following scores: 2, 4, 3, and 7 ($M = 4$).
5. Explain the difference between a definitional formula and a computational formula.
6. What is the difference between the formula for the variance you learned in this chapter and the formula that is typically used to figure the variance in research articles?

Answers

1. (a) The variance is the average of the squared deviations of each score from the mean. (b) The variance tells you about how spread out the scores are (that is, their variability), while the mean tells you the representative value of the distribution.
2. (a) The standard deviation is the square root of the average of the squared deviations from the mean. (b) The standard deviation is the square root of the variance. (c) The standard deviation tells you approximately the average amount that scores differ from the mean.
3. $SD^2 = [\Sigma(X - M)^2]/N$. SD^2 is the variance. Σ means the sum of what follows. X stands for the scores for the variable being studied. M is the mean of the scores. N is the number of scores.
4. (a) Variance: $SD^2 = [\Sigma(X - M)^2]/N = [(2 - 4)^2 + (4 - 4)^2 + (3 - 4)^2 + (7 - 4)^2]/4$
   $= 14/4 = 3.50.$
   (b) Standard deviation: $SD = \sqrt{SD^2} = \sqrt{3.50} = 1.87.$
5. A definitional formula is the standard formula in a straightforward form that shows the meaning of what the formula is figuring. A computational formula is a mathematically equivalent variation of the definitional formula that is easier to use if figuring by hand with a lot of scores, but it tends not to show the underlying meaning.
6. The formula for the variance in this chapter divides the sum of squared deviations by N (the number of scores). The variance in research articles is usually figured by dividing the sum of squared deviations by $N - 1$ (one less than the number of scores).

Z Scores

So far in this chapter you have learned about describing a group of scores in terms of a representative score and variation. In this section, you learn how to describe a particular score in terms of where it fits into the overall group of scores. That is, you learn how to use the mean and standard deviation to create a Z score; a Z score describes a score in terms of how much it is above or below the average.

Suppose you are told that a student, Jerome, is asked the question "To what extent are you a morning person?" Jerome responds with a 5 on a 7-point scale, where $1 =$ *not at all* and $7 =$ *extremely*. Now suppose that we do not know anything about how other students answer this question. In this situation, it is hard to tell whether Jerome is more

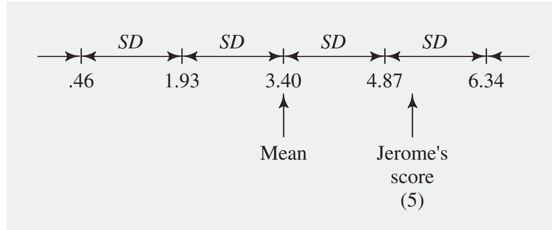


Figure 10 Score of one student, Jerome, in relation to the overall distribution on the measure of the extent to which students are morning people.

or less of a morning person in relation to other students. However, suppose that we know for students in general, the mean rating (*M*) is 3.40 and the standard deviation (*SD*) is 1.47. (These values are the actual mean and standard deviation that we found for this question in a large sample of students in statistics classes at eight different universities across the United States and Canada.) With this knowledge, we can see that Jerome is more of a morning person than is typical among students. We can also see that Jerome is above the average (1.60 units more than average, that is $5 - 3.40 = 1.60$) by a bit more than students typically vary from the average (that is, students typically vary by about 1.47, the standard deviation). This is all shown in Figure 10.

What Is a *Z* Score?

A *Z* score makes use of the mean and standard deviation to describe a particular score. Specifically, a **Z score** is the number of standard deviations the actual score is above or below the mean. If the actual score is above the mean, the *Z* score is positive. If the actual score is below the mean, the *Z* score is negative. The standard deviation now becomes a kind of yardstick, a unit of measure in its own right.

In our example, Jerome has a score of 5, which is 1.60 units above the mean of 3.40. One standard deviation is 1.47 units; so Jerome's score is a little more than 1 standard deviation above the mean. To be precise, Jerome's *Z* score is +1.09 (that is, his score of 5 is 1.09 standard deviations above the mean). Another student, Ashley, has a score of 2. Her score is 1.40 units below the mean. Therefore, her score is a little less than 1 standard deviation below the mean (a *Z* score of −.95). So, Ashley's score is below the average by about as much as students typically vary from the average.

Z scores have many practical uses. They also are part of many statistical procedures. It is important that you become very familiar with them.

***Z* score** Number of standard deviations a score is above (or below, if it is negative) the mean of its distribution; it is thus an ordinary score transformed so that it better describes that score's location in a distribution.

raw score Ordinary score (or any other number in a distribution before it has been made into a *Z* score or otherwise transformed).

Z Scores as a Scale

Figure 11 shows a scale of *Z* scores lined up against a scale of raw scores for our example of the degree to which students are morning people. A **raw score** is an ordinary score as opposed to a *Z* score. The two scales are something like a ruler with inches lined up on one side and centimeters on the other.

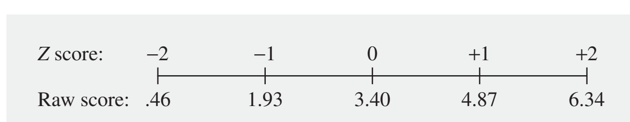


Figure 11 Scales of *Z* scores and raw scores for the example of the extent to which students are morning people.

BOX 1 The Psychology of Statistics and the Tyranny of the Mean

Looking in the behavioral and social science research journals, you would think that statistical methods are their sole tool and language, but there have also been rebellions against the reign of statistics. We are most familiar with this issue in psychology, where one of the most unexpected oppositions came from the leader of behaviorism, the school in the history of psychology most dedicated to keeping the field strictly scientific.

Behaviorism opposed the study of inner states because inner events are impossible to observe objectively. (Today most research psychologists attempt to measure inner events indirectly but objectively.) Behaviorism's most famous advocate, B. F. Skinner, was quite opposed to statistics. He was constantly pointing to the information lost by averaging the results of a number of cases. For instance, Skinner (1956) cited the example of three overeating mice—one naturally obese, one poisoned with gold, and one whose hypothalamus had been altered. Each had a different curve for learning to press a bar for food. If these learning curves had been merged statistically, the result would have represented no actual eating habits of any real mouse at all. As Skinner said, "These three individual curves contain more information than could probably ever be generated with measures requiring statistical treatment" (p. 232).

A different voice of caution was raised by another school in the history of psychology, humanistic psychology, which began in the 1950s as a "third force" in reaction to Freudian psychoanalysis and behaviorism. The point of humanistic psychology was that human consciousness should be studied intact, as a whole, as it is experienced by individuals, and it can never be fully explained by reducing it to numbers (any more than it can be reduced to words). Each individual's experience is unique.

Today, the rebellion is led in psychology by qualitative research methodologies (e.g., Willig & Stainton-Rogers, 2008), an approach that is much more prominent in other behavioral and social sciences, such as communication. The qualitative research methods developed mainly in cultural anthropology can involve long interviews or observations of a few individuals. The mind of the researcher is the main tool because, according to this approach, only that mind can find the important relationships among the many categories of events arising in the respondent's speech. Many who favor qualitative methods argue for a blend: First, discover the important categories through a qualitative approach; then, determine their incidence in the larger population through quantitative methods.

Finally, Carl Jung, founder of Jungian psychology, sometimes spoke of the "statistical mood" and its effect on a person's feeling of uniqueness. Jungian analyst Marie Louise von Franz (1979) wrote about Jung's thoughts on this subject: When we walk down a street and observe the hundreds of blank faces and begin to feel diminished, or even so overwhelmed by overpopulation that we are glad that humans don't live forever, this is the statistical mood. Yet, there is at least as much irregularity to life as ordinariness. As she puts it,

> The fact that this table does not levitate, but remains where it is, is only because the billions and billions and billions of electrons which constitute the table tend statistically to behave like that. But each electron in itself could do something else. (pp. IV–17)

Likewise, when we are in love, we feel that the other person is unique and wonderful. Yet in a statistical mood, we realize that the other person is ordinary, like many others.

Jung did not cherish individual uniqueness just to be romantic about it, however. He held that the important contributions to culture tend to come from people thinking at least a little independently or creatively, and their independence is damaged by this statistical mood.

Furthermore, von Franz (1979) argues that a statistical mood is damaging to love and life. In particular, feeling the importance of our single action makes immoral acts—war and killing, for example—less possible. We cannot count the dead as numbers but must treat them as persons with emotions and purposes, like ourselves.

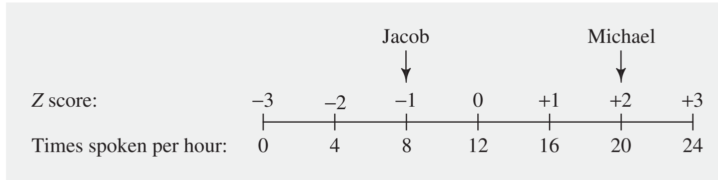


Figure 12 Number of times each hour that two children spoke, shown as raw scores and Z scores.

Suppose that a developmental specialist observed 3-year-old Jacob in a standardized laboratory situation playing with other children of the same age. During the observation, the specialist counted the number of times Jacob spoke to the other children. The result, over several observations, is that Jacob spoke to other children about eight times per hour of play. Without any standard of comparison, it would be hard to draw any conclusions from this. Let's assume, however, that it was known from previous research that under similar conditions the mean number of times children speak is 12, with a standard deviation of 4. Clearly, Jacob spoke less often than other children in general, but not extremely less often. Jacob would have a Z score of -1 ($M = 12$ and $SD = 4$, thus a score of 8 is 1 SD below M), as shown in Figure 12.

Suppose Michael was observed speaking to other children 20 times in an hour. Michael would clearly be unusually talkative, with a Z score of $+2$ (see Figure 12). Michael speaks not merely more than the average but more by twice as much as children tend to vary from the average!

Z Scores as Providing a Generalized Standard of Comparison

Another advantage of Z scores is that scores on completely different variables can be made into Z scores and compared. With Z scores, the mean is always 0 and the standard deviation is always 1. Suppose the same children in our example were also measured on a test of language skill. In this situation, we could directly compare the Z scores on language skill to the Z scores on speaking to other children. Let's say Jacob had a score of 100 on the language skill test. If the mean on that test was 82 and the standard deviation was 6, then Jacob is much better than average at language skill, with a Z score of $+3$. Thus, it seems unlikely that Jacob's less-than-usual amount of speaking to other children is due to poorer-than-usual language skill (see Figure 13).

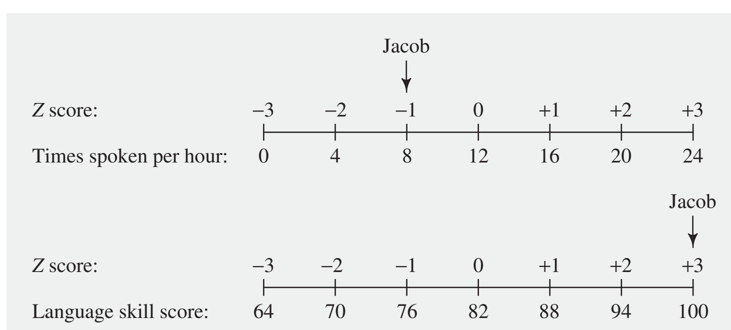


Figure 13 Scales of Z scores and raw scores for number of times spoken per hour and language skill, showing the first child's score on each.

Notice in this latest example that by using *Z* scores, we can directly compare the results of both the specialist's observation of the amount of talking and the language skill test. This is almost as wonderful as being able to compare apples and oranges! Converting a number to a *Z* score is a bit like converting the words for measurement in various obscure languages into one language that everyone can understand—inches, cubits, and zinqles (we made up that last one), for example, into centimeters. It is a very valuable tool.

Formula to Change a Raw Score to a *Z* Score

As we have seen, a *Z* score is the number of standard deviations by which the raw score is above or below the mean. To figure a *Z* score, subtract the mean from the raw score, giving the deviation score. Then, divide the deviation score by the standard deviation. The formula is

$$Z = \frac{X - M}{SD} \tag{5}$$

A *Z* score is the raw score minus the mean, divided by the standard deviation.

For example, using the formula for Jacob, the child who spoke to other children eight times in an hour (where the mean number of times children speak is 12 and the standard deviation is 4),

$$Z = \frac{8 - 12}{4} = \frac{-4}{4} = -1.$$

Steps to Change a Raw Score to a *Z* Score

❶ **Figure the deviation score: subtract the mean from the raw score.**
❷ **Figure the *Z* score: divide the deviation score by the standard deviation.**

Using these steps for Jacob, the child who spoke with other children eight times in an hour,

❶ **Figure the deviation score: subtract the mean from the raw score.** $8 - 12 = -4$.
❷ **Figure the *Z* score: divide the deviation score by the standard deviation.** $-4/4 = -1$.

Formula to Change a *Z* Score to a Raw Score

To change a *Z* score to a raw score, the process is reversed: multiply the *Z* score by the standard deviation and then add the mean. The formula is

$$X = (Z)(SD) + M \tag{6}$$

The raw score is the *Z* score multiplied by the standard deviation, plus the mean.

Suppose a child has a *Z* score of 1.5 on the number of times spoken with another child during an hour. This child is 1.5 standard deviations above the mean. Because the standard deviation in this example is 4 raw score units (times spoken), the child is 6 raw score units above the mean. The mean is 12. Thus, 6 units above the mean is 18. Using the formula,

$$X = (Z)(SD) + M = (1.5)(4) + 12 = 6 + 12 = 18.$$

Steps to Change a *Z* Score to a Raw Score

❶ **Figure the deviation score: multiply the *Z* score by the standard deviation.**
❷ **Figure the raw score: add the mean to the deviation score.**

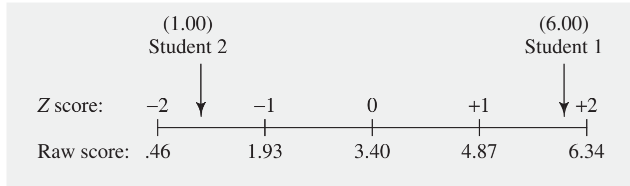


Figure 14 Scales of Z scores and raw scores for the example of the extent to which students are morning people, showing the scores of two sample students.

Using these steps for the child with a Z score of 1.5 on the number of times spoken with another child during an hour:

❶ **Figure the deviation score: multiply the Z score by the standard deviation.** $1.5 \times 4 = 6$.

❷ **Figure the raw score: add the mean to the deviation score.** $6 + 12 = 18$.

Additional Examples of Changing Z Scores to Raw Scores and Vice Versa

Consider again the example in which students were asked the extent to which they were a morning person. Using a scale from 1 (*not at all*) to 7 (*extremely*), the mean was 3.40 and the standard deviation was 1.47. Suppose a student's raw score is 6. That student is well above the mean. Specifically, using the formula,

$$Z = \frac{X - M}{SD} = \frac{6 - 3.40}{1.47} = \frac{2.60}{1.47} = 1.77.$$

That is, the student's raw score is 1.77 standard deviations above the mean (see Figure 14, Student 1). Using the 7-point scale (from $1 =$ *not at all* to $7 =$ *extremely*), to what extent are *you* a morning person? Now figure the Z score for your rating (your raw score).

Another student has a Z score of -1.63, a score well below the mean. (This student is much less of a morning person than is typically the case for students.) You can find the exact raw score for this student using the formula,

$$X = (Z)(SD) + M = (-1.63)(1.47) + 3.40 = -2.40 + 3.40 = 1.00.$$

That is, the student's raw score is 1.00 (see Figure 14, Student 2).

Let's also consider some examples from the study of students' stress ratings. The mean stress rating of the 30 statistics students (using a 0–10 scale) was 6.43, and the standard deviation was 2.56. Figure 15 shows the raw score and Z score scales.

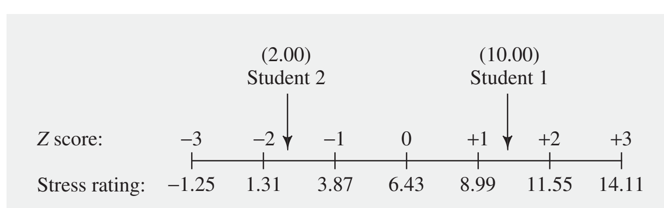


Figure 15 Scales of Z scores and raw scores for 30 statistics students' ratings of their stress level, showing the scores of two sample students. (Data based on Aron et al., 1995.)

Suppose a student's stress raw score is 10. That student is well above the mean. Specifically, using the formula,

$$Z = \frac{X - M}{SD} = \frac{10 - 6.43}{2.56} = \frac{3.57}{2.56} = 1.39.$$

That is, the student's stress level is 1.39 standard deviations above the mean (see Figure 15, Student 1). On a scale of 0–10, how stressed have *you* been in the last 2½ weeks? Figure the *Z* score for your raw stress score.

Another student has a *Z* score of -1.73, a stress level well below the mean. You can find the exact raw stress score for this student using the formula,

$$X = (Z)(SD) + M = (-1.73)(2.56) + 6.43 = -4.43 + 6.43 = 2.00.$$

That is, the student's raw stress score is 2.00 (see Figure 15, Student 2).

How are you doing?

1. How is a *Z* score related to a raw score?
2. Write the formula for changing a raw score to a *Z* score and define each of the symbols.
3. For a particular group of scores, $M = 20$ and $SD = 5$. Give the *Z* score for (a) 30, (b) 15, (c) 20, and (d) 22.5.
4. Write the formula for changing a *Z* score to a raw score and define each of the symbols.
5. For a particular group of scores, $M = 10$ and $SD = 2$. Give the raw score for a *Z* score of (a) +2, (b) +.5, (c) 0, and (d) −3.
6. Suppose a person has a *Z* score for overall health of +2 and a *Z* score for overall sense of humor of +1. What does it mean to say that this person is healthier than she is funny?

Answers

1. A *Z* score is the number of standard deviations a raw score is above or below the mean.
2. $Z = (X - M)/SD$. *Z* is the *Z* score; *X* is the raw score; *M* is the mean; *SD* is the standard deviation.
3. (a) $Z = (X - M)/SD = (30 - 20)/5 = 10/5 = 2$; (b) −1; (c) 0; (d) +.5.
4. $X = (Z)(SD) + M$. *X* is the raw score; *Z* is the *Z* score; *SD* is the standard deviation; *M* is the mean.
5. (a) $X = (Z)(SD) + M = (2)(2) + 10 = 4 + 10 = 14$; (b) 11; (c) 10; (d) 4.
6. This person is more above the average in health (in terms of how much people typically vary from average in health) than she is above the average in humor (in terms of how much people typically vary from the average in humor).

Mean, Variance, Standard Deviation, and *Z* Scores in Research Articles

The mean and the standard deviation (and occasionally, the variance) are commonly reported in research articles. The median and the mode are less often reported in research articles. *Z* scores are rarely reported in research articles. Sometimes the mean and standard deviation are included in the text of an article. For example, our fictional political scientist, in a research article about the mayors of this region, would write, "At the time of the study, the mean number of years in office for the 10 mayors in this region was 6.00 $(SD = 2.57)$." Means and standard deviations are also often listed in tables, especially if a study includes several groups or several different variables. For example, Selwyn (2007) conducted a study of gender-related perceptions of information and communication technologies (such as games machines, DVD players, and cell phones). The researchers asked 406 college students in Wales to rate eight technologies in terms of their level of masculinity or femininity. The students rated each technology using a 7-point response scale, from -3 for very feminine to $+3$ for very masculine, with a midpoint of 0 for neither masculine or feminine. Table 3 (reproduced from Selwyn's article) shows the mean, standard deviation, and variance of the students' ratings of each technology. As the table shows, games machines were rated as being more masculine than feminine, and landline telephones were rated as being slightly more feminine than masculine. Also notice that there was more variation in the ratings of landline telephones $(SD = 1.03)$ than in the ratings of television sets $(SD = 0.78)$. Notice that Table 3 is one of those rare examples where the variance is shown (usually just the standard deviation is given). Overall, the table provides a useful summary of the descriptive results of the study.

Another interesting example is shown in Table 4 (reproduced from Norcross, Kohout, & Wicherski, 2005). The table shows the application and enrollment statistics for psychology doctoral programs, broken down by area of psychology and year (1973, 1979, 1992, and 2003). The table does not give standard deviations, but it does give both means and medians. For example, in 2003 the mean number of applicants to doctoral counseling psychology programs was 71.0, but the median was only 59. This suggests that some programs had very high numbers of applicants that skewed the distribution. In fact, you can see from the table that for almost every kind of program and for both applications and enrollments, the means are typically higher than the medians. You may also be struck by just how competitive it is to get into doctoral programs in many areas of psychology. It is our experience that one of the factors that makes a lot of difference is doing well in statistics courses!

Table 3 Mean Scores for Each Technology

	N	Mean	S.D.	Variance
Games machine (e.g., Playstation)	403	1.92	1.00	.98
DVD Player	406	.44	.85	.73
Personal Computer (PC)	400	.36	.82	.68
Digital radio (DAB)	399	.34	.99	.98
Television set	406	.26	.78	.62
Radio	404	−.01	.81	.65
Mobile phone	399	−.19	.88	.77
Landline telephone	404	−.77	1.03	1.07

Note: Mean scores range from -3 (*very feminine*) to $+3$ (*very masculine*). The midpoint score of .0 denotes "neither masculine nor feminine."

Source: Selwyn, N. (2007). Hi-tech = guy-tech? An exploration of undergraduate students' gendered perceptions of information and communication technologies. *Sex Roles, 56*, 525–536.

Table 4 Application and Enrollment Statistics by Area and Year: Doctoral Programs

	N of programs				Applications								Enrollments			
					M				*Mdn*				*M*		*Mdn*	
Program	**1973**	**1979**	**1992**	**2003**	**1973**	**1979**	**1992**	**2003**	**1973**	**1979**	**1992**	**2003**	**1992**	**2003**	**1992**	**2003**
Clinical	105	130	225	216	314.4	252.6	191.1	142.0	290	234	168	126	12.0	15.4	8	8
Clinical neuro				20				72.3				37		10.7		6
Community	4	2	5	13	90.5		24.4	23.5	60		23	21	3.2	3.3	2	3
Counseling	29	43	62	66	133.4	90.9	120.2	71.0	120	84	110	59	7.3	6.8	6	7
Health			7	13			40.7	71.2			30	56	4.4	6.7	5	4
School	30	39	56	57	78.5	54.0	31.3	38.7	53	34	32	31	5.4	6.9	5	5
Other health service provider subfield				52				83.5				48		9.2		7
Cognitive			47	104			24.6	30.1			22	22	2.6	3.4	2	3
Developmental	56	72	97	111	54.1	38.9	27.6	25.5	41	30	24	22	2.8	3.4	2	2
Educational	23	28	30	35	67.8	39.7	20.0	19.7	34	26	12	13	6.0	4.9	4	4
Experimental	118	127	78	40	56.2	33.2	31.3	26.7	42	25	26	17	4.4	4.1	3	3
I/O	20	25	49	60	39.9	54.7	66.2	46.9	37	48	70	41	4.9	4.7	4	4
Neuroscience				53				22.0				16		2.8		2
Personality	23	15	10	18	42.5	24.7	12.3	47.8	33	17	6	31	1.0	2.8	1	2
Psychobiological/physiological				18				21.1				17		2.4		2
Quantitative	40	43	76	17	33.2	29.3	20.0	11.2	29	24	20	11	3.9	1.9	2	1
Social	58	72	59	85	46.7	30.9	47.1	43.1	40	24	37	35	3.3	3.2	3	3
Other fields	60	47	288	101	61.6	74.1	26.6	26.0	27	25	15	17	3.3	3.8	2	3
Total	566	645	1,089	1,079	106.1	85.2	69.4	59.6			31	33	5.6	6.7	4	4

Note. The academic years correspond to the 1975–1976, 1981–1982, 1994, and 2005 editions of *Graduate Study in Psychology,* respectively. Clinical neuro = clinical neuropsychology; I/O = industrial-organizational.
Source: Norcross, J. C., Kohout, J. L., & Wicherski, M. (2005). Graduate study in psychology: 1971–2004. *American Psychologist, 60,* 959–975. Copyright © 2005 by the American Psychological Association. Reproduced with permission. The use of APA information does not imply endorsement by APA.

Learning Aids

Summary

1. The mean is the most commonly used way of describing the representative value of a group of scores. The mean is the ordinary average—the sum of the scores divided by the number of scores. In symbols, $M = (\sum X)/N$.
2. Other, less frequently used ways of describing the representative value of a group of scores are the mode (the most common single value) and the median (the value of the middle score if all the scores were lined up from lowest to highest).
3. The variability of a group of scores can be described by the variance and the standard deviation.
4. The variance is the average of the squared deviations of each score from the mean. In symbols, $SD^2 = [\sum(X - M)^2]/N$.
5. The standard deviation is the square root of the variance. In symbols, $SD = \sqrt{SD^2}$. It is approximately the average amount that scores differ from the mean.
6. A Z score is the number of standard deviations that a raw score is above or below the mean. Among other uses, with Z scores you can compare scores on variables that have different scales.
7. Means and standard deviations are often given in research articles in the text or in tables. Z scores are rarely reported in research articles.

mean (M)
$\sum$ (sum of)
X (scores in the distribution of the variable X)
N (the number of scores)
mode
median
outlier
variance (SD^2)
deviation score
squared deviation score
sum of squared deviations
standard deviation (SD)
computational formula
definitional formula
Z score
raw score

Figuring the Mean

Find the mean for the following scores: 8, 6, 6, 9, 6, 5, 6, 2.

Answer

You can figure the mean using the formula or the steps.

Using the formula: $M = (\sum X)/N = 48/8 = 6$.

Using the steps:

❶ **Add up all the scores.** $8 + 6 + 6 + 9 + 6 + 5 + 6 + 2 = 48$.
❷ **Divide this sum by the number of scores.** $48/8 = 6$.

Finding the Median

Find the median for the following scores: 1, 7, 4, 2, 3, 6, 2, 9, 7.

Answer

❶ **Line up all the scores from lowest to highest.** 1, 2, 2, 3, 4, 6, 7, 7, 9.
❷ **Figure how many scores there are to the middle score by adding 1 to the number of scores and dividing by 2.** There are nine scores, so the middle score is 9 plus 1, divided by 2, which comes out to 5. The middle score is the fifth score.
❸ **Count up to the middle score or scores.** The fifth score is a 4, so the median is 4.

Figuring the Sum of Squared Deviations and the Variance

Find the sum of squared deviations and the variance for the following scores: 8, 6, 6, 9, 6, 5, 6, 2. (These are the same scores used in the example for the mean: $M = 6$.)

Answer

You can figure the sum of squared deviations and the variance using the formulas or the steps.

Using the formulas:

$$\begin{aligned}\text{Sum of squared deviations} &= \Sigma(X - M)^2 = (8-6)^2 + (6-6)^2 \\ &\quad + (6-6)^2 + (9-6)^2 + (6-6)^2 \\ &\quad + (5-6)^2 + (6-6)^2 + (2-6)^2 \\ &= 2^2 + 0^2 + 0^2 + 3^2 + 0^2 + -1^2 + 0^2 + -4^2 \\ &= 4 + 0 + 0 + 9 + 0 + 1 + 0 + 16 \\ &= 30.\end{aligned}$$

$$SD^2 = [\Sigma(X - M)^2]/N = 30/8 = 3.75.$$

Table 5 shows the figuring, using the following steps:

Table 5 Figuring for Example Worked-Out Problem for the Sum of Squared Deviations and Variance Using Steps

Score	Mean	❶ Deviation	❷ Squared Deviation
8	6	2	4
6	6	0	0
6	6	0	0
9	6	3	9
6	6	0	0
5	6	−1	1
6	6	0	0
2	6	−4	16
			$\Sigma = 30$ ❸

❹ Variance = 30/8 = 3.75

❶ **Subtract the mean from each score.** This gives deviation scores of 2, 0, 0, 3, 0, −1, 0, −4.
❷ **Square each of these deviation scores.** This gives squared deviation scores of 4, 0, 0, 9, 0, 1, 0, 16.
❸ **Add up the squared deviation scores.** $4 + 0 + 0 + 9 + 0 + 1 + 0 + 16 = 30$. This is the sum of squared deviations.
❹ **Divide the sum of squared deviations by the number of scores.** The sum of squared deviations, 30, divided by the number of scores, 8, gives a variance of 3.75.

Figuring the Standard Deviation

Find the standard deviation for the following scores: 8, 6, 6, 9, 6, 5, 6, 2. (These are the same scores used in the example for the mean, sum of squared deviations, and variance. $SD^2 = 3.75$.)

Answer

You can figure the standard deviation using the formula or the steps.

Using the formula: $SD = \sqrt{SD^2} = \sqrt{3.75} = 1.94$.

Using the steps:

❶ **Figure the variance.** The variance (from above) is 3.75.
❷ **Take the square root.** The square root of 3.75 is 1.94.

Outline for Writing Essays on Finding the Mean, Variance, and Standard Deviation

1. Explain that the mean is a type of representative value of a group of scores. Mention that the mean is the ordinary average, the sum of the scores divided by the number of scores.
2. Explain that the variance and standard deviation both measure the amount of variability (or spread) among a group of scores.
3. The variance is the average of each score's squared difference from the mean. Describe the steps for figuring the variance.
4. Roughly speaking, the standard deviation is the average amount that scores differ from the mean. Explain that the standard deviation is directly related to the variance and is figured by taking the square root of the variance.

Changing a Raw Score to a *Z* Score

A distribution has a mean of 80 and a standard deviation of 20. Find the *Z* score for a raw score of 65.

Answer

You can change a raw score to a *Z* score using the formula or the steps.

Using the formula: $Z = (X - M)/SD = (65 - 80)/20 = -15/20 = -.75$.

Using the steps:

❶ **Figure the deviation score: subtract the mean from the raw score.** $65 - 80 = -15$.
❷ **Figure the *Z* score: divide the deviation score by the standard deviation.** $-15/20 = -.75$.

Changing a Z Score to a Raw Score

A distribution has a mean of 200 and a standard deviation of 50. A person has a Z score of 1.26. What is the person's raw score?

Answer

You can change a Z score to a raw score using the formula or the steps.

Using the formula: $X = (Z)(SD) + M = (1.26)(50) + 200 = 63 + 200 = 263$.

Using the steps:

❶ **Figure the deviation score: multiply the Z score by the standard deviation.** $1.26 \times 50 = 63$.

❷ **Figure the raw score: add the mean to the deviation score.** $63 + 200 = 263$.

Outline for Writing Essays Involving Z Scores

1. If required by the problem, explain the mean, variance, and standard deviation as shown earlier.
2. Describe the basic idea of a Z score as a way of describing where a particular score fits into an overall group of scores. Specifically, a Z score shows the number of standard deviations a score is above or below the mean.
3. Explain the steps for figuring a Z score from a raw score (an ordinary score).
4. Mention that changing raw scores to Z scores puts scores on different variables onto the same scale, which makes it easier to make comparisons between the scores on the variables.

These problems involve figuring. Most real-life statistics problems are done on a computer with special statistical software. Even if you have such software, do these problems by hand to ingrain the method in your mind. To learn how to use a computer to solve statistics problems like those in this chapter, refer to the "Using SPSS" section at the end of this chapter.

All data are fictional unless an actual citation is given.

Set I (for answers, see the end of this chapter)

1. For the following scores, find the (a) mean, (b) median, (c) sum of squared deviations, (d) variance, and (e) standard deviation:

   32, 28, 24, 28, 28, 31, 35, 29, 26.

2. For the following scores, find the (a) mean, (b) median, (c) sum of squared deviations, (d) variance, and (e) standard deviation:

   6, 1, 4, 2, 3, 4, 6, 6.

3. Here are the noon temperatures (in degrees Celsius) in a particular Canadian city on Thanksgiving Day for the 10 years from 2000 through 2009: 0, 3, 6, 8, 2, 9, 7, 6, 4, 5. Describe the representative (typical) temperature and the amount of

variation to a person who has never had a course in statistics. Give three ways of describing the representative temperature and two ways of describing its variation, explaining the differences and how you figured each. (You will learn more if you try to write your own answer first, before reading our answer at the back of the chapter.)

4. A researcher is studying the amygdala (a part of the brain involved in emotion). Six participants in a particular fMRI (brain scan) study are measured for the increase in activation of their amygdala while they are viewing pictures of violent scenes. The activation increases are .43, .32, .64, .21, .29, and .51. Figure the (a) mean and (b) standard deviation for these six activation increases. (c) Explain what you have done and what the results mean to a person who has never had a course in statistics.
5. On a measure of concern for the environment, the mean is 79 and the standard deviation is 12 (scores on the measure can vary from 0 to 150). What are the Z scores for each of the following raw scores? (a) 91, (b) 68, and (c) 103.
6. If the mean of a measure is -11.46 and the standard deviation is 2.28, what are the Z scores for each of the following raw scores? (a) -13.12, (b) -7.26, and (c) -11.23.
7. If the mean of a measure is 145 and the standard deviation is 19, what are the raw scores for each of the following Z scores? (a) 0, (b) 1.43, (c) -2.54.
8. Six months after a divorce, the former wife and husband each take a test that measures divorce adjustment. The wife's score is 63, and the husband's score is 59. Overall, the mean score for divorced women on this test is 60 $(SD = 6)$; the mean score for divorced men is 55 $(SD = 4)$. Which of the two has adjusted better to the divorce in relation to other divorced people of their own gender? Explain your answer to a person who has never had a course in statistics.
9. A researcher studied the number of nights students reported having too little sleep over a 4-week period. In an article describing the results of the study, the researcher reports: "The mean number of nights of too little sleep was 6.84 $(SD = 3.18)$." Explain these results to a person who has never had a course in statistics.
10. In a study by Gonzaga, Keltner, Londahl, and Smith (2001), romantic couples answered questions about how much they loved their partner and also were videotaped while revealing something about themselves to their partner. The videotapes were later rated by trained judges for various signs of affiliation. Table 6 shows some of the results. Explain to a person who has never had a course in statistics the results for self-reported love for the partner and for the number of seconds "leaning toward the partner."

Set II

11. (a) Describe and explain the difference between the mean, median, and mode. (b) Make up an example (not in the text or in your lectures) in which the median would be the preferred measure of the representative value of a group of scores.
12. (a) Describe the variance and standard deviation. (b) Explain why the standard deviation is more often used as a descriptive statistic than the variance.
13. For the following scores, find the (a) mean, (b) median, (c) sum of squared deviations, (d) variance, and (e) standard deviation:

    2, 2, 0, 5, 1, 4, 1, 3, 0, 0, 1, 4, 4, 0, 1, 4, 3, 4, 2, 1, 0.

Table 6 Mean Levels of Emotions and Cue Display in Study 1

	Women (n = 60)		Men (n = 60)	
Indicator	***M***	***SD***	***M***	***SD***
Emotion reports				
Self-reported love	5.02	2.16	5.11	2.08
Partner-estimated love	4.85	2.13	4.58	2.20
Affiliation-cue display				
Affirmative head nods	1.28	2.89	1.21	1.91
Duchenne smiles	4.45	5.24	5.78	5.59
Leaning toward partner	32.27	20.36	31.36	21.08
Gesticulation	0.13	0.40	0.25	0.77

Note: Emotions are rated on a scale of 0 (*none*) to 8 (*extreme*). Cue displays are shown as mean seconds displayed per 60 s.
Source: Gonzaga, G. C., Keltner, D., Londahl, E. A., & Smith, M. D. (2001). Love and the commitment problem in romantic relations and friendship. *Journal of Personality and Social Psychology, 81,* 247–262. Copyright © 2001 by The American Psychological Association. Reproduced with permission. The use of APA information does not imply endorsement by APA.

14. For the following scores, find the (a) mean, (b) median, (c) sum of squared deviations, (d) variance, and (e) standard deviation:

    1,112; 1,245; 1,361; 1,372; 1,472.

15. For the following scores, find the (a) mean, (b) median, (c) sum of squared deviations, (d) variance, and (e) standard deviation:

    3.0, 3.4, 2.6, 3.3, 3.5, 3.2.

16. For the following scores, find the (a) mean, (b) median, (c) sum of squared deviations, (d) variance, and (e) standard deviation:

    8, −5, 7, −10, 5.

17. Make up three sets of scores: (a) one with the mean greater than the median, (b) one with the median and the mean the same, and (c) one with the mode greater than the median. (Each made-up set of scores should include at least five scores.)
18. A researcher interested in political behavior measured the square footage of the desks in the official office of four U.S. governors and of four chief executive officers (CEOs) of major U.S. corporations. The figures for the governors were 44, 36, 52, and 40 square feet. The figures for the CEOs were 32, 60, 48, and 36 square feet. (a) Figure the mean and the standard deviation for the governors and for the CEOs. (b) Explain what you have done to a person who has never had a course in statistics. (c) Note the ways in which the means and standard deviations differ, and speculate on the possible meaning of these differences, presuming that they are representative of U.S. governors and large U.S. corporations' CEOs in general.
19. A developmental specialist studies the number of words seven infants have learned at a particular age. The numbers are 10, 12, 8, 0, 3, 40, and 18. Figure the (a) mean, (b) median, and (c) standard deviation for the number of words learned by these seven infants. (d) Explain what you have done and what the results mean to a person who has never had a course in statistics.
20. On a measure of artistic ability, the mean for college students in New Zealand is 150 and the standard deviation is 25. Give the Z scores for New Zealand college students who score (a) 100, (b) 120, (c) 140, and (d) 160. Give the raw scores for persons whose Z scores on this test are (e) −1, (f) −.8, (g) −.2, and (h) +1.38.

21. On a standard measure of peer influence among adolescents, the mean is 300 and the standard deviation is 20. Give the *Z* scores for adolescents who score (a) 340, (b) 310, and (c) 260. Give the raw scores for adolescents whose *Z* scores on this measure are (d) 2.4, (e) 1.5, (f) 0, and (g) −4.5.
22. A person scores 81 on a test of verbal ability and 6.4 on a test of math ability. For the verbal ability test, the mean for people in general is 50 and the standard deviation is 20. For the math ability test, the mean for people in general is 0 and the standard deviation is 5. Which is this person's stronger ability, verbal or math? Explain your answer to a person who has never had a course in statistics.
23. A study involves measuring the number of days absent from work for 216 employees of a large company during the preceding year. As part of the results, the researcher reports, "The number of days absent during the preceding year $(M = 9.21; SD = 7.34)$ was . . . ." Explain the material in parentheses to a person who has never had a course in statistics.
24. Goidel and Langley (1995) studied the positivity and negativity of newspaper accounts of economic events in the period just before the 1992 U.S. Presidential election. Table 7, reproduced from their report, describes the numbers of front-page articles on economic news in the *New York Times* for the 23 months preceding the election. Explain the results in the Mean and Standard Deviation columns to a person who has never had a course in statistics. (Be sure to explain some specific numbers as well as the general principle.)

Table 7 Descriptive Statistics for News Coverage Variables Aggregated by Month. *New York Times Index,* January 1981–November 1992.

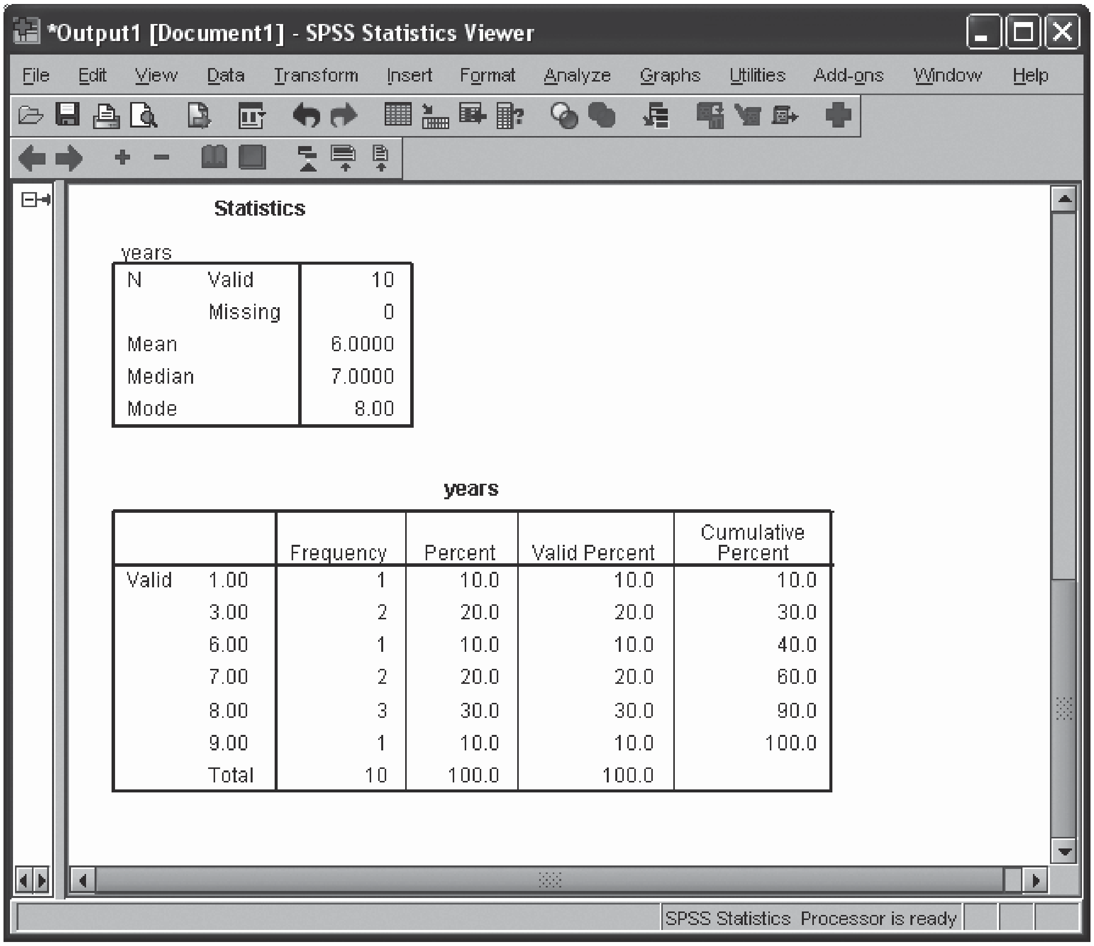
	Mean	Standard Deviation	Range	Total
Total Front-Page Articles	5.84	4.10	0–22	835
Positive Front-Page Articles	1.64	1.33	0–6	261
Negative Front-Page Articles	1.83	1.92	0–11	234

Source: New York Times Index.
From Goidel, R. K., & Langley. R. E. (1995). Media coverage of the economy and aggregate economic evaluations: Uncovering evidence of indirect media effects. *Political Research Quarterly, 48,* 313–328. Copyright © 1995 by the University of Utah. Reprinted by permission of Sage Publications, Inc.

The ✐ in the following steps indicates a mouse click. (We used SPSS version 17.0 to carry out these analyses. The steps and output may be slightly different for other versions of SPSS.)

Finding the Mean, Mode, and Median

❶ Enter the scores from your distribution in one column of the data window.
❷ ✐ *Analyze.*
❸ ✐ *Descriptive statistics.*
❹ ✐ *Frequencies.*
❺ ✐ the variable for which you want to find the mean, mode, and median, and then ✐ the arrow.



Statistics

years

N	Valid	10
	Missing	0
Mean		6.0000
Median		7.0000
Mode		8.00

years

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1	10.0	10.0	10.0
	3.00	2	20.0	20.0	30.0
	6.00	1	10.0	10.0	40.0
	7.00	2	20.0	20.0	60.0
	8.00	3	30.0	30.0	90.0
	9.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

Figure 16 Using SPSS to find the mean, median, and mode for the example of the number of years served in office by city mayors.

6 *Statistics.*
7 *Mean,* *Median,* *Mode,* *Continue.*
8 Optional: To instruct SPSS *not* to produce a frequency table, the box labeled *Display frequency tables* (this *un*checks the box).
9 *OK.*

Practice the preceding steps by finding the mean, mode, and median for the mayors' example at the start of the chapter. (The scores in that example are: 7, 8, 8, 7, 3, 1, 6, 9, 3, 8.) Your output window should look like Figure 16. (If you instructed SPSS not to show the frequency table, your output will show only the mean, median, and mode.)

Finding the Variance and Standard Deviation

As mentioned earlier in the chapter, most calculators and computer software—including SPSS—calculate the variance and standard deviation using a formula that involves dividing by $N - 1$ instead of N. So, if you request the variance and standard deviation directly from SPSS (for example, by clicking *variance* and *std. deviation* in Step 7 above), the answers provided by SPSS will be different than the answers in this chapter.[2] The following steps show you how to use SPSS to figure the variance and standard deviation using the dividing-by-N method you learned in this chapter. It is easier to learn these steps using actual numbers, so we will use the mayors' example again.

❶ Enter the scores from your distribution in one column of the data window (the scores are 7, 8, 8, 7, 3, 1, 6, 9, 3, 8). We will call this variable "years."
❷ Find the mean of the scores by following the preceding steps for Finding the Mean, Mode, and Median. The mean of the years variable is 6.
❸ You are now going to create a new variable that shows each score's squared deviation from the mean. *Transform*, *Compute Variable.* You can call the new variable any name that you want, but we will call it "squared_deviation." So, write *squared_deviation* in the box labeled *Target Variable*. You are now going to tell SPSS how to figure this new variable called squared_deviation. In the box labeled *Numeric Expression*, write $(years - 6) * (years - 6)$. (The asterisk is how you show *multiply* in SPSS.) As you can see, this formula takes each score's deviation

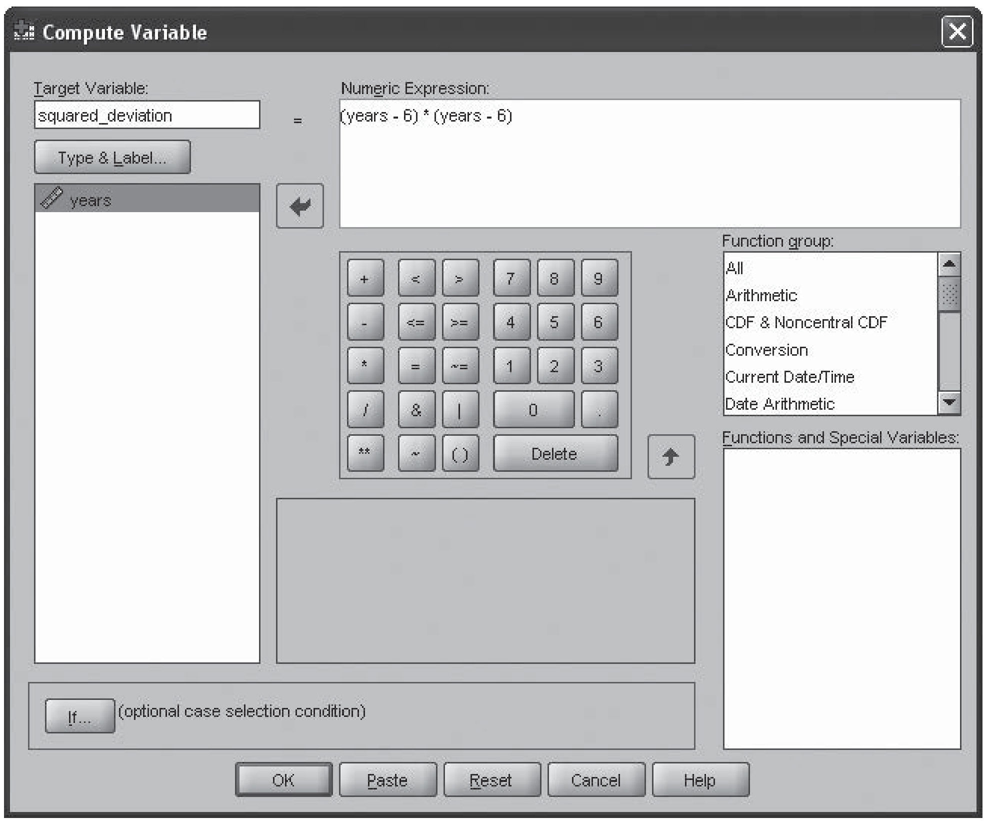


Figure 17 SPSS compute variable window for Step ❸ of finding the variance and standard deviation for the example of the number of years served in office by city mayors.

[2]Note that if you request the variance from SPSS, you can convert it to the variance as we figure it in this chapter by multiplying the variance from SPSS by $N - 1$ (that is, the number of scores minus 1) and then dividing the result by N (the number of scores). Taking the square root of the resulting value will give you the standard deviation (using the formula you learned in this chapter). We use a slightly longer approach to figuring the variance and standard deviation in order to show you how to create new variables in SPSS.

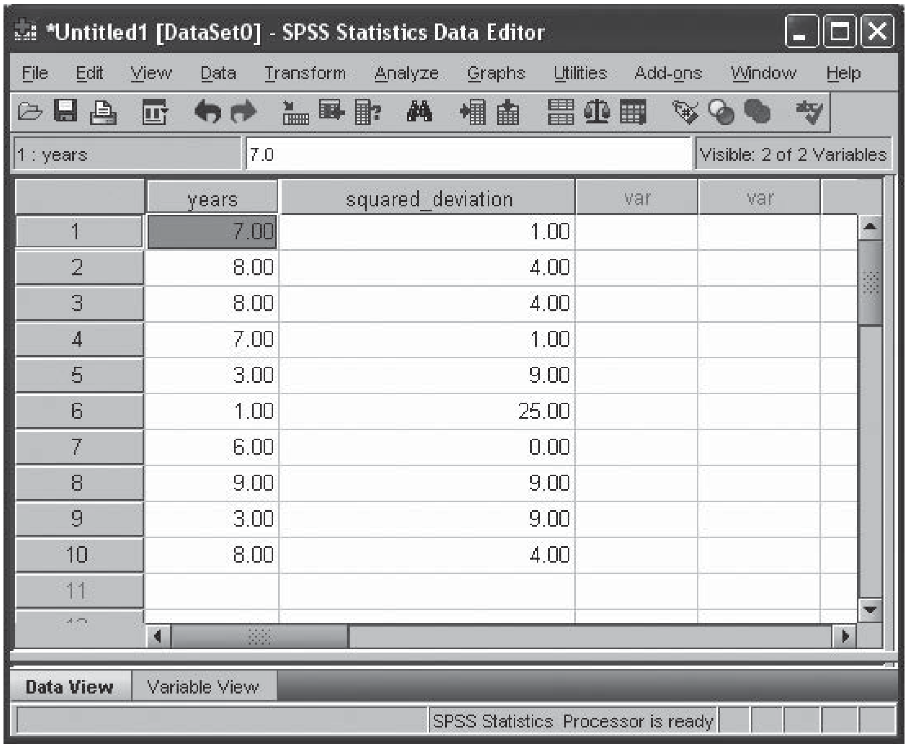


Figure 18 SPSS data window after Step ❸ of finding the variance and standard deviation for the example of the number of years served in office by city mayors.

score and multiplies it by itself to give the squared deviation score. Your Compute Variable window should look like Figure 17. *OK*. You will see that a new variable called *squared_deviation* has been added to the data window (see Figure 18). The scores are the squared deviations of each score from the mean.

❹ As you learned in this chapter, the variance is figured by dividing the sum of the squared deviations by the number of scores. This is the same as taking the mean of the squared deviation scores. So, to find the variance of the years scores, follow the steps shown earlier to find the mean of the *squared_deviation* variable. This comes out to 6.60, so the variance of the years scores is 6.60.

❺ To find the standard deviation, use a calculator to find the square root of 6.60, which is 2.57.

If you were conducting an actual research study, you would most likely request the variance and standard deviation directly from SPSS. However, for our purposes in this chapter (describing the variation in a group of scores), the steps we just outlined are entirely appropriate. Also, following this procedure will further engrain the principles in your mind, and also teaches you how to create new variables in SPSS.

Changing Raw Scores to *Z* Scores

It is easier to learn these steps using actual numbers, so we will use the mayors' example again.

❶ Enter the scores from your distribution in one column of the data window (the scores are 7, 8, 8, 7, 3, 1, 6, 9, 3, 8). We will call this variable "years."

❷ Find the mean and standard deviation of the scores using the steps above for Finding the Mean, Mode, and Median, and Finding the Variance and Standard Deviation. The mean is 6 and the standard deviation is 2.57.

❸ You are now going to create a new variable that shows the *Z* score for each raw score. *Transform*, *Compute Variable.* You can call the new variable any name that you want, but we will call it "zyears." So, write *zyears* in the box labeled *Target Variable.* In the box labeled *Numeric Expression*, write $(years - 6)/2.57$. As you can see, this formula creates a deviation score (by subtracting the mean from the raw score) and divides the deviation score by the standard deviation. *OK.* You will see that a new variable called *zyears* has been added to the data window. The scores for this *zyears* variable are the *Z* scores for the years variable. Your data window should now look like Figure 19.[3]

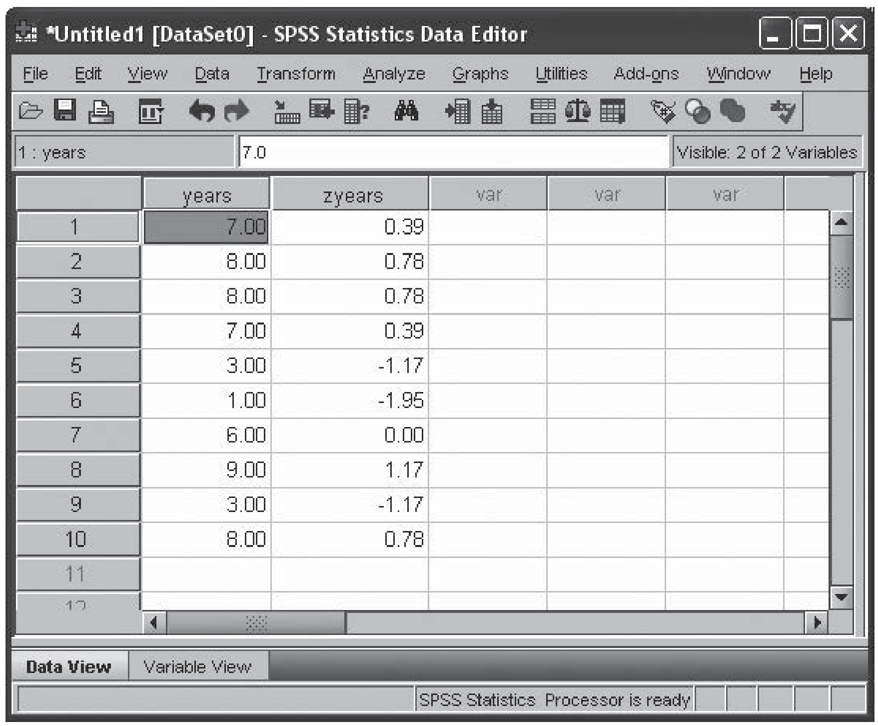


Figure 19 Using SPSS to change raw scores to *Z* scores for the example of the number of years served in office by city mayors.

[3]You can also request the *Z* scores directly from SPSS. However, SPSS figures the standard deviation based on the dividing by $N - 1$ formula for the variance. Thus, the *Z* scores figured directly by SPSS will be different from the *Z* scores as you learned to figure them. Here are the steps for figuring *Z* scores directly from SPSS: ❶ Enter the scores from your distribution in one column of the data window. ❷ *Analyze*, *Descriptive statistics*, *Descriptives*. ❸ on the variable for which you want to find the *Z* scores, and then the arrow. ❹ the box labeled *Save standardized values as variables* (this checks the box). ❺ *OK*. A new variable is added to the data window. The values for this variable are the *Z* scores for your variable. (You can ignore the output window, which by default will show descriptive statistics for your variable.)

Answers to Set I Practice Problems

1. (a) $M = (\Sigma X)/N = 261/9 = 29$; (b) 28; (c) $\Sigma(X - M)^2 = (32 - 29)^2 + (28 - 29)^2 + (24 - 29)^2 + (28 - 29)^2 + (28 - 29)^2 + (31 - 29)^2 + (35 - 29)^2 + (29 - 29)^2 + (26 - 29)^2 = 86$; (d) $SD^2 = [\Sigma(X - M)^2]/N = 86/9 = 9.56$; (e) $SD = \sqrt{SD^2} = \sqrt{9.56} = 3.09$.
2. (a) 4; (b) 4; (c) 26; (d) 3.25; (e) 1.80.
3. The average temperature, in the sense of adding up the 10 readings and dividing by 10, was 5 degrees Celsius. This is the *mean* (the ordinary average). However, if you line up the temperatures from lowest to highest, the middle two numbers are 5 and 6. Thus, the *median* is 5.5 degrees Celsius. The specific temperature that came up most often is the *mode*—the mode is 6 degrees Celsius.

   As for the variation (the amount of variability), one approach is the *variance*—the average of each temperature's squared deviation from the mean temperature, which is 7.0. You get a more direct sense of how much a group of numbers vary among themselves if you take the square root of the variance, which gives the standard deviation—the square root of 7.0 is 2.65. This means, roughly, that on an average day the temperature differs by 2.65 degrees from the average of 5 degrees.
4. (a) .40; (b) .14; (c) similar to 3 above.
5. (a) $Z = (X - M)/SD = (91 - 79)/12 = 1$; (b) $-.92$; (c) 2.
6. (a) $Z = (X - M)/SD = (-13.12 + 11.46)/2.28 = -.73$; (b) 1.84; (c) .10.
7. (a) $X = (Z)(SD) + M = (0)(19) + 145 = 145$; (b) 172.17; (c) 96.74.
8. Wife: $Z = (X - M)/SD = (63 - 60)/6 = .5$. Husband: $Z = (59 - 55)/4 = 1$. The husband has adjusted better in relation to other divorced men than the wife has adjusted in relation to other divorced women.

   For wives, a score of 63 is 3 points better than the average of 60 for divorced women in general. (The "mean" in the problem is the ordinary average—the sum of the scores divided by the number of scores.) There is, of course, some variation in scores among divorced women. The approximate average amount that women's scores differ from the average is 6 points—this is the *SD* (standard deviation) referred to in the problem. (*SD* is only approximately the average amount scores differ from the mean. To be precise, *SD* is the square root of the average of the square of the difference of each score from the mean.) Thus, the wife's score is only half as far as above the mean of wives' scores. This gives her what is called a Z score of $+.5$, which gives her location on a scale that compares her score to that of divorced women in general. Using the same logic, the husband's divorce adjustment is as much above the mean as the average amount that men differ from the mean; that is, he has a Z score of $+1$. What this all means is that both have adjusted better than average for their gender, but the husband had adjusted better in relation to other divorced men than the wife has adjusted in relation to other divorced women. Using Z scores allowed us to put the raw scores for men and women onto the same scale, which made it easier to compare their scores.
9. The "mean" is the ordinary arithmetic average—add up the total number of nights of too little sleep and divide by the number of students. The average number of nights of too little sleep over the 4-week period was 6.84. The *SD* (standard deviation) describes the variability of the scores. Roughly speaking, the standard deviation is the average amount that the number of nights of too little sleep are spread out from their average—in this case, by 3.18 nights. This is quite a lot of spread. To be more precise, you figure the standard deviation by taking each person's number of nights of too little sleep and subtracting 6.84 from it and squaring this difference; the standard deviation is the square root of the average of these differences.
10. Like the answer to 9 above, focusing on means of 5.02, 5.11, 32.27, and 31.36 and on standard deviations of 2.16, 2.08, 20.36, and 21.08.

This page intentionally left blank