Describing the Placement of Observations

1. Do you want to know an observation's location relative to the *mean?*
- If *yes,* then convert the raw score to a z score.
- If *no,* then go to step 2.

2. Do you want to know the score's standing relative to other *scores?*
- If *yes,* then convert the raw score to z score and determine its *percentile rank.*
- If *no,* then go to step 3.

3. Do you want to summarize the entire *distribution of scores?*
- If *yes,* then select measures of *central tendency* and *variability* (see chapter 4)
- If *no,* then review your goals and go back to step 1.

Converting a Raw Score to a z Score

1. Is the raw score from a *population?*
- If *yes,* then go to step 3.
- If *no,* then go to step 2.

2. Is the raw score from a *sample?*
- If *yes,* then go to step 4.
- If *no,* then go to step 7.

3. Do you know the μ and the σ of the *population?*
- If *yes,* then go to step 5.
- If *no,* determine these *population parameters* and then to go to step 5.

4. Do you know the $\bar{X}$ and the s of the *sample?*
- If *yes,* then go to step 6.
- If *no,* then determine these *sample statistics* and then go to step 6.

5. Enter the appropriate numbers into the following formula and then report the z score's location relative to the population μ: $z = (X - \mu)/\sigma$

6. Enter the appropriate numbers into the following formula and then report the z score's location relative to the sample $\bar{X}$: $z = (X - \bar{X})/s$

7. A z score cannot be calculated unless information about the raw score's parent distribution is available.

Determining the Percentile Rank of a z Score

1. Is the z score *above* the mean of 0.0?
- If *yes,* then go to step 3.
- If *no,* then go to step 2.

2. Is the z score *below* the mean of 0.0?
- If *yes,* then go to step 4.
- If *no,* then go to step 5.

3. Locate the z score in column 4 of Table B.2 in Appendix B, then:
- a. Find the area between the mean and z by reading across to column B;
- b. Convert the column B proportion to a percentage by multiplying it by 100;
- c. Add the percentage to 50%—the resulting (total) percentage is the *percentile rank* of the z score.

4. Locate the z score in column A of Table B.2 in Appendix B, then:
- a. Find the area below the z by reading across to column C.
- b. Convert the column C proportion to a percentage by multiplying it by 100.
- c. The resulting percentage is the *percentile rank* of the z score.

5. Is the z score equal to 0.0?
- a. If *yes,* then the *percentile rank* of the mean is 50%.
- b. If *no,* you have made an error; go back to step 1.

CHAPTER 5

STANDARD SCORES AND THE NORMAL DISTRIBUTION

Chapter Outline

One of the great lessons taught by research methods, statistics, and life is the importance of asking a particular question: Compared to what? The methodology used by behavioral scientists reminds them to constantly compare the behavior of one group with that of another, usually a control group. Comparing what research participants do within the various conditions of an experiment is reinforced when the data are analyzed. As we learned in chapter 4, most statistics are meaningless unless they compare the measures of central tendency and dispersion observed in one group with those drawn from others. We will broaden our conceptual understanding of the process of drawing meaning from descriptive statistics by starting to compare them in this chapter. Actual, inferential statistics will be presented later in the book, but their underlying foundations begin in earnest here.

But wait a moment, what about life—how does life teach us to ask about or to make various comparisons? The act of comparing ourselves to others is ubiquitous and ongoing in the social world. In the language of social psychology, for example, the act of evaluating our own opinions and abilities against those of other people is called social comparison (Festinger, 1954). We do it all the time, no doubt because other people—their acts, abilities, feelings—are such a good source of information for us. In fact, we cannot seem to help ourselves, as we almost effortlessly, seemingly automatically, look to others for information about what we should do or how we should feel in a particular situation.

There is no more obvious realm for comparing ourselves to others than the classroom. Take a moment and think about how often you spend time comparing your academic performance to that of others, whether real (e.g., roommate, friends, family, classmates) or imagined (e.g., the "perfect" student, the "struggling" peer). There is a good chance that your social comparing in the educational realm has been a minor preoccupation since middle or high school, if not earlier (in my experience, even the most laid back college student has a competitive streak, one prompted by comparing the self to others).

Here is a past example of your drive to compare, one that may still be a source of mild pride or subtle discomfort for your conscious self: What were your scores on the Scholastic Aptitude Test (SAT)? Way back when you received two main scores, one for verbal ability and one for math. Despite the fact that the College Board and the Educational Testing Service cautioned against adding these subtest scores together, your primary concern was presumably whether their combination put you in the "admit range" of the college or university of your choice. After that, you probably shared (and compared!) your scores—and the hopes and dreams you had for those scores—with friends and peers who had similar plans for their futures.

When examining actual behavior or descriptive statistics referring to it, be a critical observer by asking, "Compared to what?"

Back in chapter 3, we established that you were probably unaware that the percentile rank information accompanying your SAT scores provided a relative sense of how your scores compared to those of other people who took the test at the same time you did. If your verbal score was at the 70th percentile, for instance, then you performed better than or equal to 70% of predominantly high-school-aged peers who took the test. But there was also some other information that you probably did not know or really attend to—that the test was standardized so that individuals evaluating performance on the SAT would have a sense of where a given verbal or math score fell along the distribution of possible scores. In other words, people evaluating your scores were able to ask—and to answer—compared to what?

The SAT is a single but important component of student applications to college, one that has been used for admissions purposes for over 70 years, and not without its share of controversy (Schwartz, 1999; for a related discussion, see Bowen & Bok, 1998). The verbal and mathematical subtests each can have scores ranging between 200 and 800 or, if you prefer, combined scores ranging from 400 to 1,600 (but see Data Box 5.B and the end-of-chapter *Project Exercise*). The mean of each subtest was set at 500: scores falling above this point were deemed above average and scores falling below were said to be below average. Different educational institutions determine different admissions standards for acceptable test scores (i.e., how many students within a given range of scores would be considered for admission at the college or university).

Your understanding of central tendency (mean scores on the SAT), dispersion (range of possible values on the two subtests or the combined score), and percentile rank (relative standing of a given score within the distribution of SAT scores) should—albeit with some hindsight—render your own SAT scores a bit more meaningful. What remains to be explained, however, is how scores on the SAT or any other standardized test, psychological inventory, or personality measure can be compared to past, present, and yes, even *future* test takers. How are such comparisons, really extrapolations, possible? We will touch on these and related issues throughout the chapter.

To do so, we will apply knowledge acquired in the first four chapters of the book to examine what are called standard or z scores and their relation to the normal distribution. Both z scores and the normal distribution will provide us with a meaningful context for understanding how a given score or statistic can be interpreted in light of the distribution it was drawn from, an excellent preparation for the inferential statistics to come. To begin, however, we must discuss the matter of standardizing measures.

DATA BOX 5.A

Social Comparison Among Behavioral and Natural Scientists: How Many Peers Review Research Before Publication?

A hallmark of the scholarship and publication in the behavioral and natural sciences is peer review. Manuscripts—primarily journal articles but some books, as well—are anonymously reviewed by peers before they are deemed worthy of publication. Many authors routinely ask peers who work in the subfield of their discipline to read and comment on their work before it is submitted to a journal editor or a publisher. Suls and Fletcher (1983) wondered to what extent researchers from physics, chemistry, psychology, and sociology were apt to consult with colleagues in advance of submitting work for publication.

Following the work of philosopher of science Thoman Kuhn (1970), Suls and Fletcher (1983) argued that behavioral scientists would be more likely to seek colleagues' counsel about the contents of their research than would natural scientists. Kuhn persuasively suggested that the natural sciences are "mature" disciplines—they have agreed on theories, research methodologies and codified knowledge—relative to the behavioral sciences, which still seek intellectual unity. As a result, behavioral scientists deal with a higher degree of uncertainty in their work than do natural scientists and such uncertainty promotes a need to communicate and validate one's views with those of others. Suls and Fletcher posited that this difference would manifest itself rather obviously in publications—behavioral scientists would exhibit greater social comparison by acknowledging or thanking a higher average number of peers in published articles than would natural scientists.

Using an archival procedure—that is, examining, coding, and analyzing preexisting records—Suls and Fletcher (1983) counted the number of colleagues cited in footnotes, author notes, or other acknowledgment sections of published physics ($n = 220$), chemistry ($n = 209$), psychology ($n = 115$), and sociology ($n = 89$) articles (more studies from the former two disciplines were included because many more articles are published in the natural than behavioral sciences). The results indicated that natural scientists tended to cite the same number of peers, which, as predicted, fell below the number cited by either the psychologists or the sociologists. In turn, sociologists tended to thank more peers than the psychologists (see the table shown below). Further, natural science articles tended to have more authors relative to those published in the behavioral sciences, a fact that emphasizes both the collaborative nature and shared perspective of physics and chemistry (see the table shown below).

What can we conclude from these data? Behavioral scientists are more motivated to compare their scholarship with their colleagues prior to the usual round of peer review, and the certainty of knowledge in one's discipline plays a part in the process. Natural scientists appear to be more certain of the knowledge promulgated by their fields, and psychologists seem somewhat more certain of disciplinary conclusions than their colleagues in sociology. Following Suls and Fletcher (1983), a good question to ask is whether this higher level of consultation among behavioral scientists will lead to less uncertainty and, in the long run, disciplinary maturity?

Mean Number of Acknowledgements and Authors by Scientific Discipline

Variable	Physics	Chemistry	Psychology	Sociology
N acknowledged[a]	0.55	.52	1.24	2.07
N of authors[b]	3.16	2.79	1.98	1.62

Note: [a]Higher numbers reflect more peers acknowledged per article; [b]higher numbers indicate more authors per article.

Source: Adapted from Suls & Fletcher (1983).

DATA BOX 5.B

Explaining the Decline in SAT Scores: Lay Versus Statistical Accounts

Pundits have noticed a steady but modest decline in the SAT verbal and math scores for years, puzzling over the meaning and significance of the changes in these averages. Are students becoming dumb or dumber? Or, has the SAT been "dumbed down"? If these two possibilities are both false, how else can the noted decline in SAT scores be explained? To answer this question, we need to examine the lay or nonstatistical explanation for the decline, as well as a more data-driven, statistical account for the change in scores.

Lay explanations for the decline in SAT scores focus on social themes that can quickly become politicized. One popular explanation is that today's students are less intellectual or motivated than the students of past generations, having been raised on television and rock videos rather than books and newspapers. A companion theme of this "anti-intellectual" youth argument is that our public schools, teachers as well as administrators, are doing a worse job at educating students than they did 20 or 30 years ago (i.e., "in the good old days"). This latter argument tends to appear in the context of test scores or whenever school taxes are levied or increases are contemplated. You have probably seen national and local politicians of various stripes make political hay with these sorts of arguments.

Stop and think for a moment: Given the educational and technological booms the United States has enjoyed for the past few decades, is it really possible that students are becoming, as it were, "dumber"? What else might be going on to make a decline in SAT scores be more apparent than real? Well, consider the fact that in 1941, 10,000 predominantly white males from the Northeastern United States took the test—recently, more than 1 million—indeed, closer to 2 million—men and women take the SAT in a given year. These men and women hail from diverse ethic, racial, and social backgrounds as well as educational experiences (College Board, 1995).

Can you guess what a statistician might say about the SAT data and the purported decline in scores? Consider what you know about sample size and populations. First, the relative size and composition of the yearly population of college bound students has grown dramatically in size. This fact might bring to mind the law of large numbers introduced in chapter 4—the apparent decline in scores may be a false one because the performance of (recent) larger samples of students taking the SAT may be more representative of the population's parameters. In other words, the "decline" in average scores is really a better reflection of the μ of the distribution of possible SAT scores.

Second, and in a related way, the population of students now taking the test is diverse—more heterogeneous—than the homogeneous student samples from decades past. More students, different subpopulations of students, taking the test should result in some shift in the test scores—the one that has occurred just happens to be in a downward direction, which may more appropriately capture or represent the population parameters of the test. The College Board and the Educational Testing Service have actually addressed this artificial decline by performing what is called a "re-centering" of the scores, a simple transformation that returns the verbal and math subscale means back to 500 each. We will examine this transformation procedure in more detail in the *Project Exercise* at the end of this chapter.

Score realignments aside, I hope that the lesson here is clear. You should always be suspicious of claims that skills are changing—for better or worse—when at the same time the population based on which those skills are assessed is also changing. Remember the lessons taught by larger samples, as well as the theme of this chapter—compared to what?

Why Standardize Measures?

The logic behind standardizing measures is really quite simple. By the term "standardize," we refer to a process whereby a given object or measure can be interpreted or used in a consistent way across time, setting, and circumstance. Standardization enables us to understand and compare similar objects or measures on the same continuum, rendering them familiar and easy to understand. When heating the oven to 350 degrees in preparation for baking a cake, for example, you do not have to worry that the oven's heat settings somehow differ from the recipe's author (that is, unless your oven is on the fritz or you live at a high altitude). Similarly, if your shoe size is 8 and your shirt size is medium, you can walk into virtually any clothing store and find shoes and shirts that fit. Our monetary system, too, is a form of standardization—a dollar in rural Nebraska is equivalent in value to one in midtown Manhattan, despite the fact it might stretch a bit farther in the former than the latter setting. More to the point, perhaps, two disparate items can be equivalent in value—say, a book and a hair dryer—because they have been standardized on the dimension of price. Oven temperature, shoe and clothing sizes, and money, then, are all standardized.

Standardization involves consistent forms of measurement and interpretation, an attempt to consider the relative position of objects on a single dimension.

In the same way, behavioral scientists often create standardized measures that produce scores that mimic the universality of heat settings or shoe sizes. Beyond the now familiar SAT example, the other common instrument to cite in this vein is the intelligence or IQ (for "intelligence quotient") test. Like the SAT, most IQ tests are composed of more than one subtest, the contents of which sometimes vary. Some entail mathematical or verbal components, but others tap into memory, the ability to assemble or to deconstruct objects, general comprehension, and gross and fine motor skills among numerous other possibilities.

Most IQ tests yield a single score—a standard score—said to reflect the performance of an individual. In fact, the standard IQ scores form a distribution possessing a population mean (μ) of 100 and a population standard deviation (σ) of 15. Thus, an individual with a measured IQ of 110 is said to score in the above average range or in the first standard deviation around the mean (i.e., $100 + 15 = 115$, the boundary of the first standard deviation above the mean; see Figure 5.1). Someone scoring an 84 would be below average, falling just inside the second standard deviation below the mean. The first standard deviation below the mean falls at a score of 85 (i.e., $100 - 15$), while the second falls between scores of 85 and 70 (i.e., $85 - 15 = 70$). As scores fall away from the mean in either

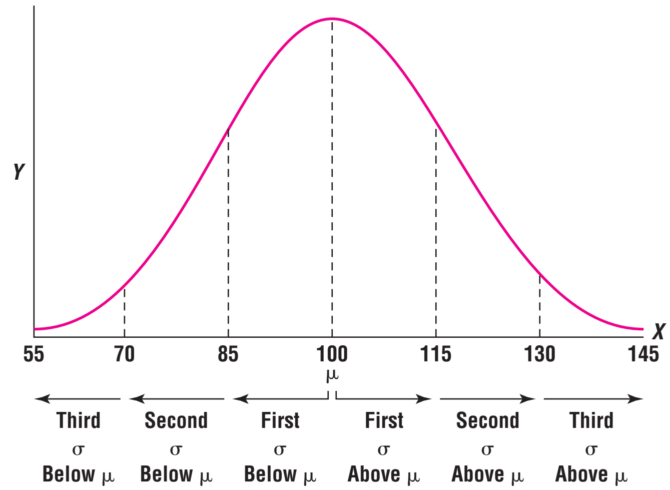


Figure 5.1 Hypothetical Distribution of Scores on an IQ Test

direction—that is, in the so-called "tails" of the distribution—they occur less frequently as compared to those scores that cluster close to the mean (note that in Figure 5.1, the median and the mode would share the population mean's value; see chapter 4). Thus, for example, a person who scores a 142 on an IQ test is certainly not unheard of, but by falling way in the distribution's upper tail, neither is she commonplace.

Various and sundry IQ tests and the SAT are by no means the only standardized tests you may know of or encounter. Other common standardized tests include the Law School Admissions Test (LSAT), the Graduate Record Exam (GRE), the Graduate Management Admissions Test (GMAT), and the Medical College Admissions Test (MCAT). There are also a host of psychological tests and inventories with unfamiliar names or acronyms that also yield standard scores.

Standardization of IQ scores or, for that matter, any measure, enables researchers to precisely locate where a particular score falls in a distribution and to describe how it compares to other scores in the distribution. This achievement involves converting a raw *score* into a *standard score.*

KEY TERM A **raw score** is any score or datum that has *not* been analyzed or otherwise transformed by a statistical procedure.

A raw score, then, is any basic score, rating, or measurement in its pure form.

KEY TERM A **standard score** is derived from a raw score. *Standard scores* report the relative placement of individual scores in a distribution and are useful for various inferential statistical procedures.

Raw scores are turned into standard scores so that they can be used for comparison purposes (e.g., Is one score closer to the mean than another?) or to make inferences (e.g., How likely is it that a given score is from one rather than another population?). When raw scores are converted into standard scores, their apparent value will change but *no* information is lost; rather, the conversion renders the score easier to work with and to compare with other scores along a distribution.

Converting raw scores to standard scores promotes comparison and inference.

In order for standard scores to be truly useful, however, the mean of the relevant distribution must also be taken into account. Does a given score fall above, at, or below the mean of its distribution? The mean continues to serve as the main reference point, the anchor, if you will, of any distribution. Any mean, of course, provides limited information unless it is accompanied by its standard deviation. We rely on the standard deviation to inform us whether a given score is similar to or divergent from the mean, as well as other scores in the distribution. For example, does a particular score fall into the first or the second standard deviation above or below a mean? To begin to answer these questions more concretely, we turn to the z score.

The *z* Score: A Conceptual Introduction

When percentile rank was presented in chapter 3, we discussed the utility of knowing what percentage of scores fell at or below a given point, a score, within a distribution. The z score, too, provides placement information about the location of a score within a distribution but, unlike percentile rank, it relies on the mean and the standard deviation. Specifically, the z score provides information about the relative position between some observed score and the mean, and it is reported in terms of relative deviation (i.e., the mean deviation is reported in standard deviation units, a topic that was introduced in chapter 4).

Let's begin with a straightforward example. Imagine that you have a score of 55 drawn from a distribution with a mean of 50. We know that the absolute difference between these numbers is 5 (i.e., $55 - 50 = 5$), that the observed score is greater than the known mean—but how much greater? In other words, we want to know the *relative*

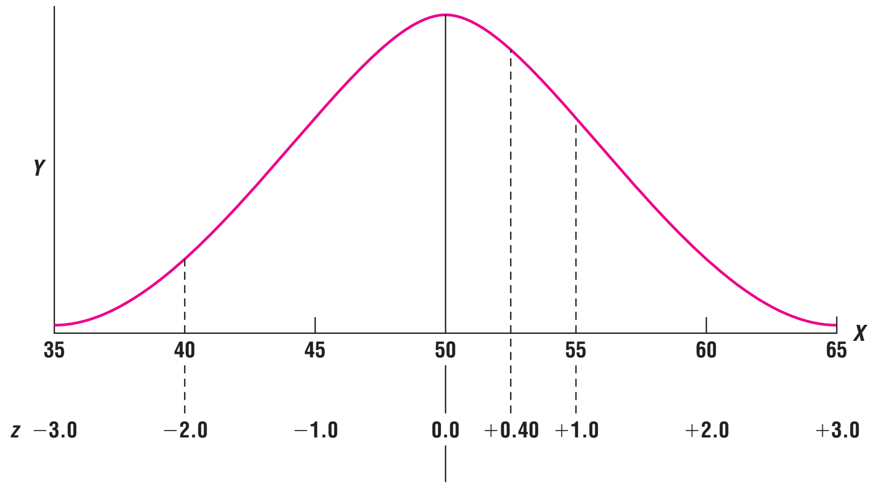


Figure 5.2 Distribution of Raw Scores and Corresponding z Scores Where $\mu = 50$ and $\sigma = 5$

difference between the mean and the observed score. Imagine, then, that we also know the standard deviation, which is equal to 5. If we divide the absolute difference between the mean and the observed score (i.e., 5) by the standard deviation of 5, we will know the relative deviation of the observed score from the mean. That is, $5/5 = +1.0$, indicating that a score of 55 lies 1.0 standard deviation above the mean of 50 (see Figure 5.2). In fact, anytime a relative deviation is positive in value, we know that the score falls above the mean—and we know where it falls in terms of standard deviation units.

What if the observed score were equal to 40? This time, although we know the absolute difference between the mean and the observed score is 10, we will have a negative number (i.e., $40 - 50 = -10.0$). If we divide -10.0 by the standard deviation of 5, we find that a score of 40 lies 2.0 standard deviations below the mean (i.e., $-10.0/5 = -2.0$; see Figure 5.2). Anytime a relative deviation is negative, then, we know how many standard deviation units below the mean it is located.

What have we accomplished? We have just calculated two *z scores,* common types of standard scores derived from raw scores.

KEY TERM

A descriptive statistic, the ***z* score** indicates the distance between some observed score (X) and the mean of a distribution in standard deviation units.

The z score tells us one very important thing: *How many standard deviations away from the mean is a given score?* As we just learned in the above examples, a score of 55 was 1.0 standard deviation above the mean, while a score of 40 happened to be 2.0 standard deviations below the mean. Put another way, we know that the first score was relatively closer to the mean than the second score (see Figure 5.2). In fact, a defining characteristic of any z distribution is that the width of the standard deviation around the mean will *always* be equal to 1.0.

The standard deviation of a z distribution is *always* 1.0.

Please note that a z score need not fall precisely on a standard deviation as the two scores in this example did. I used numbers that fell precisely on the standard deviations for convenience. If we have an observed score of 52, for example, then the corresponding z score would be equal to $+0.40$ (i.e., $52 - 50/5 = 2/5 = +0.40$), which falls less than halfway across the first standard deviation above the mean (see Figure 5.2). That is, $+0.40$ is less than $+1.0$, which tells us that a score of 52 is very close to the mean of 50, just as $+0.40$ is very close to the z distribution's mean of 0.

The mean of the *z* distribution is always equal to 0.0.

That's right—the mean of any distribution of *z* scores is always equal to 0. Students can initially be put off by the mean of the *z* distribution always being equal to 0 because they assume that means it is equal to "nothing." Don't be put off in this way: Think of a *z* score, any *z* score, as a guidepost that tells you an observed (raw) score's relative location within a distribution of scores. When you calculate a *z* of 0, it simply means that an observed score is equal to the mean, just as negative *z*s fall below the mean and positive *z*s place themselves above it.

Key Points Regarding *z* Scores. Let's review these three important characteristics of *z* scores:

1. The mean of any *z* distribution is always 0.
2. The standard deviation of the any *z* distribution is always 1.0.
3. When the value of a *z* score is positive (+), then the score falls above the mean of 0; when it is negative (−), it falls below it. The only time a *z* score lacks a sign is when it is equal to 0 (i.e., the raw score is equivalent to the original mean of the distribution).

The third point merits more emphasis and explanation. When we use the + and − signs, we again use them as guideposts for the relative placement of a score above or below the mean. We do not treat these signs as being indicative of value per se. Although it is true that a *z* of +1.50 is greater than a *z* of −1.2, it is appropriate to think of the former as being higher in value or magnitude than the latter by virtue of its location relative to the mean.

But wait, there is one more point to add to this list:

4. The distribution of *z* scores will always retain the shape of the distribution of the original raw scores.

Standardization ≠ normalization; conversion to *z* scores does not alter the shape of a distribution in any way.

Students—and many faculty members, actually—often forget this last one because they misunderstand the definition of standard scores. Specifically, they may erroneously believe that the conversion to *z* scores somehow "sanitizes" or "normalizes" the data, taking an any-old-shaped distribution of raw scores and turning it into that paragon of statistical virtue, a normal distribution (see below). Not so. The conversion to relative deviation provides a much better indication of where scores lie relative to one another, but it does *not* change the shape of the distribution in any way—any skew or kurtosis present, as well as any outlying scores, remains intact.

A Brief Digression on Drawing *z* Distributions for Plotting Scores. Despite the likelihood of an irregular shaped distribution, I have always found it useful to draw a normal shaped distribution when I work with *z* scores. Figure 5.2 is a formal example of what I have in mind, as it enables you to visualize the placement of one score relative to another, and to identify their relation to the mean of 0 and the 1-unit standard deviations surrounding it. Again, few distributions of scores you encounter will actually be normal, but jotting down a bell-shaped curve is a heuristic convenience, a short cut or rule of thumb, that makes interpreting the converted raw scores easier to conceptualize.

Your bell-shaped curve need not be elegantly drawn, perfectly symmetric, or even particularly neat to look at—but it should enable you to "see" how scores in the data relate to one another. A simple way to proceed is to sketch what I call a "volcano" shape first (see panel (a) in Figure 5.3), and then to draw a semicircle over its top (see panel (b) in Figure 5.3). Once you add this lid, you can erase the lines underneath it (see panel (c) in Figure 5.3) and draw in lines representing the mean and standard deviations (see panel (d) in Figure 5.3). It then becomes a snap to mark in the *z* score or

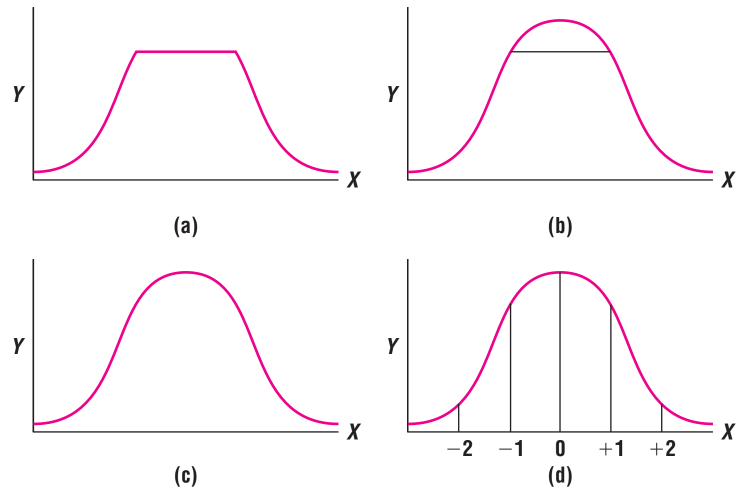


Figure 5.3 Sketching a Simple Bell-Curve to Represent a Distribution of *z* Scores

scores on the distribution for ready reference. I promise you that these quick curve drawings will come in handy in later sections of this chapter and for solving the problems at its end.

Comparing Different Measures and Distributions. There is a conceptual as well as practical advantage associated with the conversion of raw scores to *z* scores that must be mentioned here. The use of *z* scores enables researchers to compare measures from different distributions with one another, despite the fact they have different means and standard deviations. Because any different measures—besides SAT or IQ scores, self-esteem, depression, grade point average (GPA) are obvious candidates—can be converted to *z* scores, an individual's relative placement on one measure can be compared with any other (see the *Application* discussion later in this chapter). Conceptually, then, these various measures represent variables available for comparison, and instead of only comparing one person's relative level of self-esteem with his GPA, we can perform the same comparison for many people. This advantage will become increasingly important, and we will revisit it directly in the next chapter when we discuss correlational relationships among variables in detail.

Formulas for Calculating *z* Scores

In general, we talk about *z* scores in terms of populations, but they are also frequently calculated from sample data. Here is the formula used to calculate a *z* score from sample data:

[5.1.1] $$z = \frac{X - \overline{X}}{s}.$$

z Scores can be calculated from sample or population data.

As you can see, X represents the known raw score, $\overline{X}$ is the sample mean, and s is the sample's standard deviation. The formula for calculating a *z* score from population data is conceptually identical—only the symbols change:

[5.2.1] $$z = \frac{X - \mu}{\sigma}.$$

Once again, X is the known raw score, but now μ represents the mean of the population and σ is its standard deviation.

What if you know a z score and you want to transform it back to its original raw score form? This reverse conversion is really quite simple. Besides the z score itself, you must know the value of the mean and the standard deviation of the distribution of raw scores, and then it becomes a simple matter of multiplication and addition. Here is the transformation formula back to a sample's raw score:

[5.3.1] $$X = \overline{X} + z(s).$$

As shown in [5.3.1], the original raw score can be determined by multiplying the z score times the sample standard deviation, which is then added to the sample mean. Let's use sample data from earlier in the chapter to demonstrate this reconversion. Recall that we earlier calculated that a score of 55 from a distribution with a mean of 50 and a standard deviation of 5 corresponded to a z score of 1.0. If we enter the z, the mean ($\overline{X}$), and the standard deviation (s) into [5.3.1] we can show that the raw score of X is indeed 55:

[5.3.2] $$X = 50 + 1.0(5),$$

[5.3.3] $$X = 50 + 5,$$

[5.3.4] $$X = 55.$$

The transformation formula for converting a population-based z score back to its raw form is conceptually identical but symbolically different:

[5.4.1] $$X = \mu + z(\sigma).$$

In this version of the transformation formula, the population standard deviation (σ) is multiplied by the z score, the product of which is then added to the population mean (μ).

Conceptually, of course, both the sample and population formulas achieve the same end. The distance between an observed score and the mean of its distribution is divided by the standard deviation of that mean. Before we learn more about the statistical utility of z scores, we need to review the standard normal distribution, which will help us to work with the information the scores provide.

The Standard Normal Distribution

In chapter 4, the basic shape of distributions, notably the normal distribution, was introduced. You will recall that the normal distribution is bell-shaped, featuring a preponderance of observations about its middle (the center of the bell) and gradually fewer as you move in either direction away from the center area into the tails surrounding it. One half of the normal distribution is the mirror image of the other half; that is, 50% of the available observations can be found on either side. In its idealized form, the mean, median, and mode of the normal distribution will share the same values.

Most people tend to think of the normal distribution as a two-dimensional image, as a line that is drawn left to right, but one that rises up halfway into a bell shape before it begins a symmetric decline. In fact, the normal curve can be thought of in three-dimensional terms, like a symmetric half-bubble or gradual hump rising from an otherwise flat surface. Due to convenient displays, such as blackboards and texts like this one, however, we are used to thinking of the normal curve in flat terms—when, in fact, only if we were to split a three-dimensional curve in half would it appear this way. My point is that although we cannot presently look at a three-dimensional normal distribution together, we can try to think three-dimensionally about it while learning its properties.

The normal distribution was created as an abstract ideal useful to mathematicians and statisticians. These two groups used the normal curve as a way to account for numerical relationships, notably the relative likelihood of some numerical events occurring instead of others, or probability (see chapter 8). In reality, there is not one normal distribution but rather a "family" of curves that can be defined as normal (Elifson, Runyon, & Haber, 1990). The reason so many exist is the endless potential for different combinations of population means (μ) and population standard deviations (σ). As we will see, all normal curves share some basic characteristics, but they are not "cookie-cutter" constructs; some are larger than others, flatter in shape, or have a steeper peak (recall the discussion of distribution shapes in chapter 3).

Across time, researchers realized that the normal distribution had a wide variety of applications and that it was a very effective way to describe the behavior of various naturally occurring events. Natural and behavioral scientists were quick to realize that the normal curve was useful for studying phenomena germane to their respective areas of inquiry. A wide variety of sociological, psychological, and biological variables distribute themselves normally or in an approximately normal fashion, or they lend themselves to transformation to a normal curve (e.g., Rosenthal & Rosnow, 1991). The discipline of psychology, for example, would be at a loss if the properties of the normal distribution—notably its usefulness for hypothesis testing (see chapter 9)—were not available. Of course, some everyday variables are also normally distributed. Height and weight represent everyday examples of variables that qualify in this regard.

Although we will not be using it directly, there is a formula that specifies the shape of the normal distribution. Here it is:

[5.5.1]
$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}.$$

Take a good look at it—it won't bite you. Forgive me, but I merely want you to think about the following information pertaining to this formula and not be intimidated by it. The conceptual information the formula provides will be useful background material for the rest of this section of the chapter. For statisticians, the main advantage of having this formula is that although available data can change or vary, this formula can always be used to determine what a normal distribution looks like, no matter what the value of its mean and standard deviation. What we see in [5.5.1] is that the relative frequency or function of any score (X) is dependent upon the population mean (μ) and variance (σ^2), the constant π (which is $\cong$ 3.146), and the constant e (the base of the natural logarithm, which is $\cong$ 2.7183). In other words, if the relative frequencies of X were entered into the equation, we would be able to see how they must form the now familiar normal curve.

Of course, such plotting of scores is not our purpose here, because we already know—or at least assume—that the normal distribution provides us with certain statistical advantages. Chief among these advantages is the ability to partition the area under the normal curve to identify what proportion or percentage of observations must fall within given ranges.

Standard Deviation Revisited: The Area Under the Normal Curve

In some sense, normal distributions are "standardized" because particular percentages of observations occur in a predictable pattern. (For ease of comparison, we will rely on percentages, but discussing proportions under the curve is equally appropriate; recall the relation between percentages and proportions introduced in chapter 3.) When the area under the curve was described as being predictable in chapter 4, you did not know anything about z scores or greater detail about the utility of the normal distribution. We noted then that approximately 68% of the available observations fall within the first standard deviation interval above and below the mean, or 34.13% in each one. Formula [5.5.1] actually enables us to specify what percentage of observations fall within each of these standard deviation intervals occurring to the left and the right of the mean.

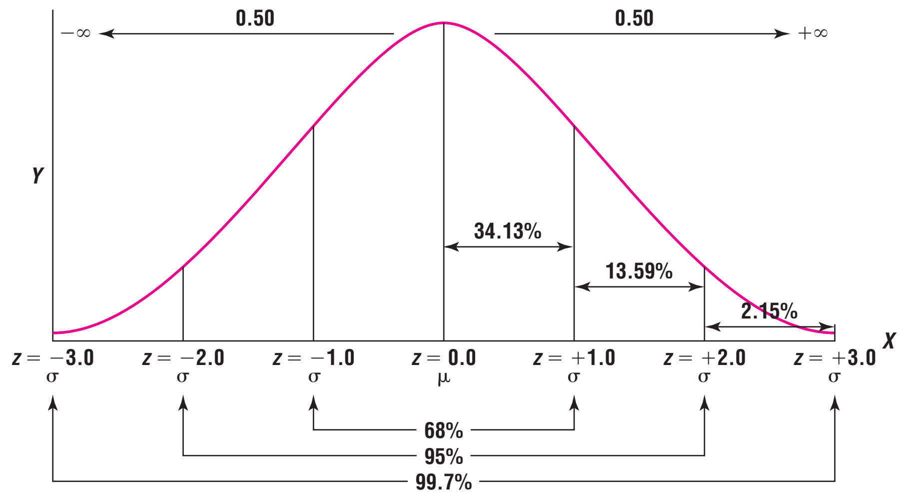


Figure 5.4 Area Between Standard Deviation Intervals Along a *z* Distribution

Our understanding of *z* scores and their fixed standard deviation interval widths of 1.0 will now pay off. Theoretically, the total area under the normal curve is equal to 100%. Figure 5.4 illustrates a standard normal distribution with *z* scores running along the *x* axis. As you can see in Figure 5.4, on either side of the mean of 0 is one standard deviation interval equal to 34.13% of the area under the normal curve (i.e., $2 \times 34.14\% = 68.26\%$, the available area under the curve). The area between the first and second standard deviation on either side of the mean is equal to 13.59% (i.e., $2 \times 13.58\% = 27.18\%$ of the available area; see Figure 5.4). In the third standard deviation from the mean resides 2.15% of observations (i.e., $2 \times 2.15\% = 4.30\%$ of the available area; see Figure 5.4). If you add the total area accounted for under the curve in Figure 5.4—you have accounted for 99.74% of the available observations or *z* scores.

What about this less than 1% of other (potential) observations? They appear to remain but are not accounted for under the curve—why? In practical terms, these observations fall relatively far in either tail, the resting place of true outliers. But there is a theoretical consideration regarding the normal curve that must be noted as well. The normal curve never "closes" because its two tails do not end (asymptotic to the x-axis)—that is, it is comprised of a potentially infinite number of cases (Elifson et al., 1990). As a result, notice that the sign for infinity—both positive ($+\infty$) and negative ($-\infty$)—is included in Figure 5.4. These signs serve as theoretical reminders about the nonclosing nature of the normal distribution, but they also have some practical significance. These signs reinforce the idea that researchers are actually trying to generalize findings to populations that are large and subject to change, but still measurable.

Contrary to expectation, a normal distribution never ends (theoretically) because its tails remain open.

Beyond the uniformity of observations falling in standard deviations, what else do we know about normal distributions? For most practical and theoretical purposes, the standard normal distribution is treated as having outer boundaries that end at ± 3.0 standard deviations (see Figure 5.4). Normal distributions are symmetric (lack any skew) and are described as generally mesokurtic (see chapter 4). Finally, using a normal distribution, the percentile ranking of any given *z* score is easily determined.

Application: Comparing Performance on More than One Measure

It may have dawned on you that because *z* scores are standard scores, they can allow researchers to compare empirical "apples," as it were, with "oranges." That is, different measures can be compared with one another even if they have different numbers of

Table 5.1 Scores on Three Hypothetical Measures of Psychological Well-Being

Measure	Raw Score	Population Parameters	*z* Score
Depression	80	$\mu = 110, \sigma = 15$	−2.00
Self-esteem	90	$\mu = 75, \sigma = 8$	+1.88
Life-satisfaction	25	$\mu = 40, \sigma = 5$	−3.00

items, as well as different means and standard deviations. The ability to covert disparate measures to comparable scores is invaluable for researchers, as it frees them to consider how all kinds of variables can affect one another.

We can consider a simple example in this vein. A clinical psychologist might be interested in examining a client's scores on a few standardized measures, say, a depression inventory, a self-esteem scale, and a life satisfaction scale. Perhaps the clinician wants to confirm her assessment that the client is not depressed, but merely dissatisfied with some life circumstances involving both home and work. The client's raw scores on the standardized measures, the population mean and sample deviation for each measure, and the *z* scores are shown in Table 5.1.

When converted to *z* scores, different, even disparate, variables can be compared to one another.

As shown in Table 5.1, the *z* score for the client's depression level ($z = -2.00$) is relatively far below the mean, which indicates a low likelihood of depression (i.e., higher scores reflect a greater incidence of depression). The *z* score corresponding to self-esteem ($z = +1.88$), however, falls fairly high above the mean (i.e., the client has a relatively high level of self-esteem). In contrast, the *z* score representing life satisfaction ($z = -3.00$) is not in the desired direction—the client is clearly dissatisfied with salient aspects of his life—as it is three standard deviations below the mean. As shown by this simple example, then, scores from different scales can be compared relative to one another once they are converted to standard scores. The psychologist can now focus on helping the client to recognize which aspects of his life need to be addressed, and the comparison of some empirical "apples" and "oranges" allowed that to happen.

Knowledge Base

1. What are the characteristics of any distribution of *z* scores?
2. You have a sample with a mean of 25 and a standard deviation of 3. What are the *z* scores corresponding to the following raw scores?
   a. 18 b. 26 c. 25 d. 32 e. 15
3. You have a population with a μ of 65 and a σ of 6. Convert the following *z* scores back to their raw score equivalents.
   a. +1.2 b. −2.5 c. −1.0 d. 0 e. +2.9
4. In percentage terms, what is the total area under the normal curve? What percentage of the observations fall in the first standard deviation below the mean?

Answers

1. Any *z* distribution has a mean of 0 and a standard deviation of 1.0. Positive *z* scores are always greater than the mean, whereas negative *z* scores are less than the mean in value. A distribution of *z* scores will retain the shape of the original raw score distribution.

2. a. −2.33 b. +0.33 c. 0 d. +2.33 e. −3.33
3. a. 72.2 b. 50 c. 59 d. 65 e. 82.4
4. 100%; 34.13%

Working with *z* Scores and the Normal Distribution

In general, working with statistics entails not only performing some calculations but also learning to work with—that is, to read—statistical tables. Earlier in the book, for example, you learned to understand how to use a table of random numbers (see the *Project Exercise* in chapter 2 and Table B.1 in Appendix B). In this section of the text, you will learn to use Table B.2 from Appendix B, which contains the approximate proportions of the area under the normal curve (for convenience, proportions will be converted to percentages). Table B.2 can be used to address questions regarding the relative placement of a given *z* score in relation to the mean (i.e., the percentage of cases existing between the mean and the *z*), to specify the percent of cases falling above the *z* score, and finally, to identify the percentage of cases occurring between two *z* scores.

Learning to use Table B.2 really serves two purposes. It will address the obvious need posed by this section of the chapter—how to locate *z* scores and identify the area they delineate under the normal curve. Working with this table will also prepare you for what is to come later, a variety of other statistical tables that can be read and understood with the same ease. Remember, statistical tables exist to make data analysis easier, not to hinder you or to create undue anxiety. Whenever I begin to use any statistical table, I always remind myself what its purpose is, while at the same time I ask myself, "What am I doing now?" That is, I make sure that I know what I am reading in the table—and why—and that I am not simply going through the motions of flipping pages and locating numbers without really understanding the rationale for doing so.

To begin, please turn to Table B.2 in Appendix B. As you can see, Table B.2 has three columns—one for *z* scores (A), one indicating the area between the mean and *z* (B), and one denoting the area beyond *z* (C). The *z* scores in column A all appear to be positive—what happened to the negative scores in the distribution? Recall that the *z* distribution is symmetric; in other words, except for the plus (+) or minus (−) signs, each side of the distribution is equal to the other. In practical terms, then, when working with a negative *z* score, we need only to attach a negative sign (−) to an entry in column A.

The entries in columns B and C are proportions. To convert them to percentages, we simply multiply them by 100 (i.e., the decimal point is moved two places to the right). A proportion of .3389, for example, would be equal to 33.89%. When rounded, this number could be reported as 34%.

Before we review several examples incorporating *z* scores and our knowledge of the normal curve, we need to be sure of our interpretive bearings regarding Table B.2. Locate the entry for a *z* of 0.0 in column A (see the top left corner of the table's first page). As we know, a *z* of 0.0 corresponds to the mean of the distribution. What percentage of the cases should fall between the mean and itself? None, of course. This intuitive fact can be confirmed by examining the proportion shown in column B, which is .0000. By looking at column C, the cases occurring in the area *beyond* the *z* of 0.0—that is, above or below the mean—is shown to be .5000, or 50%. This percentage should make sense to you, as 50% of the cases *must* fall to the right and to the left of the mean in this symmetric distribution. If the logic of these facts does not make sense to you, please stop and reread the previous chapter sections explaining *z* scores and their relation to the normal curve.

Finding Percentile Ranks with *z* Scores

Imagine that a student earns a score of 35 on a test that is normally distributed. The μ of the test is 28 and the σ is 4. Using formula [5.2.1], the *z* score for a raw score of 35 would be:

[5.6.1] $$z = \frac{35.0 - 28.0}{4.0} = \frac{7.0}{4.0} = +1.75.$$

By locating $z = +1.75$ in column A of Table B.2 and then reading across to column B, we find that the area between the mean and *z* is .4599. Thus, 45.99%, really 46%, of the available scores lie between the students score and the mean (see Figure 5.5). Because we also know that 50% of the scores fall below the mean, we can say that 95.99% (i.e., 50% + 46.99%) of the area under the curve falls at or below the student's score of 35. In other words, by using the *z* distribution, we know the percentile rank of a score of 35 on this particular test.

If you examine the entry in column C corresponding to the *z* score of +1.75, you can see that 4.01% of the cases fall above it (see Figure 5.5). In other words, about 4% of the people who took the test could have scores greater than a score of 35.

What if a second student received a score of 23 on the test? Where does this score fall? Calculate a *z* score using the now familiar formula:

[5.7.1] $$z = \frac{23.0 - 28.0}{4.0} = -\frac{5.0}{4.0} = -1.25.$$

As noted earlier, negative *z* scores share the exact same proportion (and percentage) information as positive *z* scores. Turn to Table B.2 in Appendix B and locate the $z = -1.25$ entry in column A. This time, we know that our work will take place *below* the mean of 0.0 because the *z* is negative. Column B in Table B.2 indicates that the area between the mean and $z = -1.25$ is .3944 or 39.44%. This area between the mean and *z* is also highlighted in Figure 5.6. The numerical entry in column C now gives us the percentile rank of a score of 23 on the test. How so? Column C's entry—.1056 or 10.56%—represents

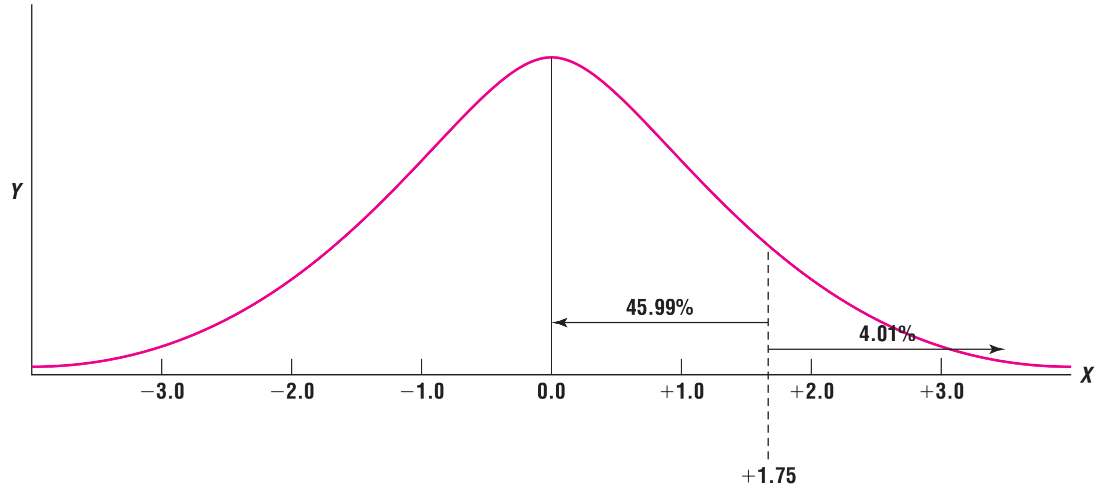


Figure 5.5 Location of $z = +1.75$ in a *z* Distribution with $\mu = 28.0$ and $\sigma = 4.0$

Note: In percentage terms, the area between the mean and the *z* score of +1.75 is 45.99%. The area in the curve beyond the *z* score of +1.75 is 4.01%.

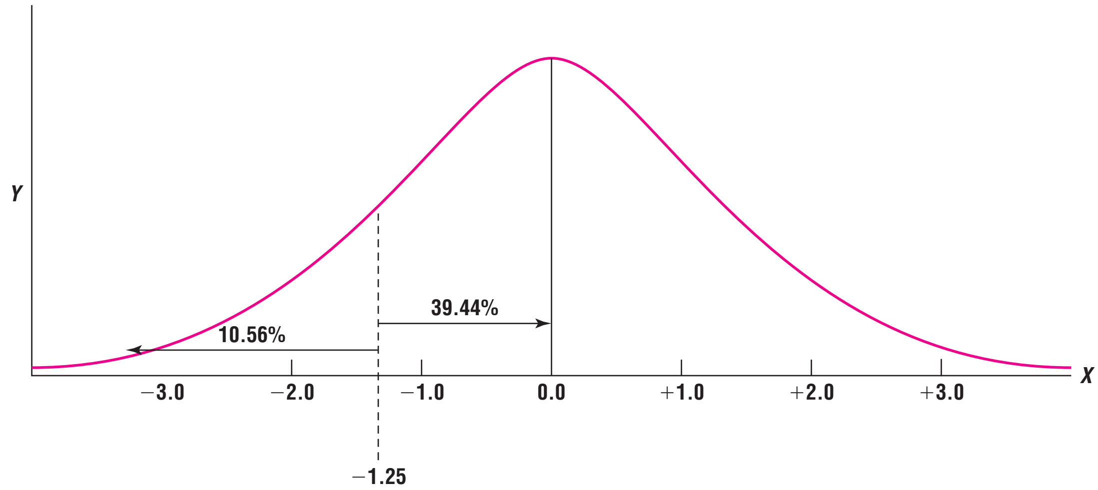


Figure 5.6 Location of $z = -1.25$ in *z* Distribution with $\mu = 28.0$ and $\sigma = 4.0$

Note: In percentage terms, the area between the mean and a *z* score of −1.25 is 39.44%. The area in the curve beyond the *z* score of −1.25 is 10.56%.

the available area of the curve that falls below $z = -1.25$ (see Figure 5.6). Thus, a score of 23 on the test fell at approximately the 11th percentile—11% of the scores fell at or below the score of 23.

Further Examples of Using *z* Scores to Identify Areas Under the Normal Curve

Besides using *z* scores to determine the percentile rank of raw scores, *z* scores can also be used to delineate the area between two raw scores. We will consider two illustrative examples, one using *z* scores falling on either side of the mean and one involving standard scores appearing on the same side of the mean.

Area Between *z* Scores on Either Side of the Mean. An investigator wants to know what percentage of the cases on a standardized instrument fall between scores of 85 and 111. The standardized test, which measures people's knowledge of geography, has a μ of 96 and a σ of 7. The first step is to convert both test scores to *z*s using formula [5.2.1]. The test score of 85 is equal to:

[5.8.1] $$z = \frac{85 - 96}{7} = -\frac{11}{7} = -1.57,$$

and the score of 111 is equal to:

[5.9.1] $$z = \frac{111 - 96}{7} = \frac{15}{7} = +2.14.$$

To begin, of course, we draw a diagram of a *z* distribution similar to the one shown in Figure 5.7. As you well know, the *z* of −1.57 is just over one and one half standard deviations below the mean and the *z* of +2.14 is slightly over the boundary of the second standard deviation to the right of the mean (see Figure 5.7). How do we proceed?

Well, because we want to know the area—the percentage of the cases—falling between scores falling on either side of the mean of 0.0, intuitively, we can (1) first determine the percentage distance between each score and the mean, and then (2) add these percentages together. Using column B in Table B.2, we learn that 44.18% of the cases fall between a *z* of −1.57 and the mean, and that 48.38% of the cases fall between a *z* of

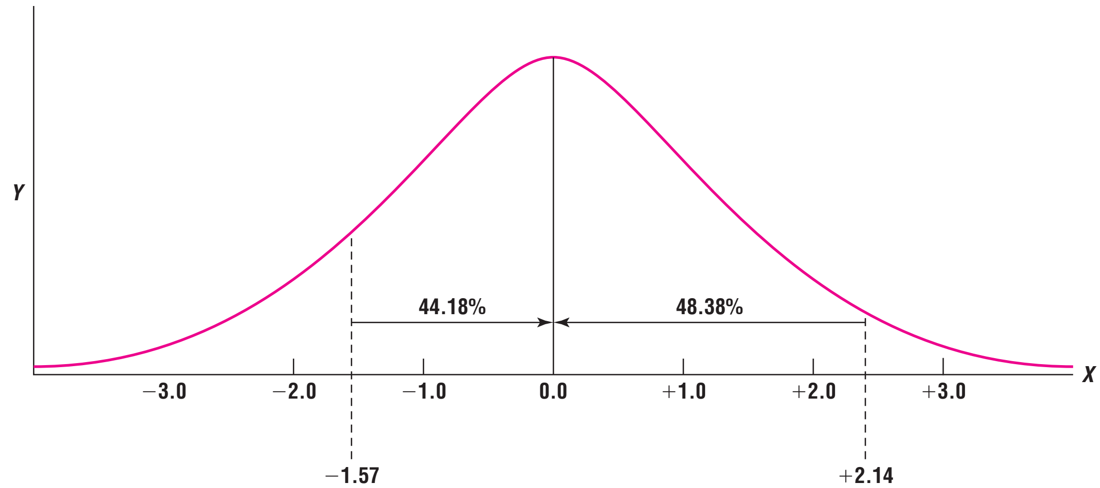


Figure 5.7 z Distribution Representing Geographic Knowledge Test with $\mu = 96$ and $\sigma = 7$

Note: The raw scores of 85.0 and 111.0 are shown as z scores of −1.57 and +2.14, respectively. The area between $z = -1.57$ and the mean is 44.18%, and between +2.14 and the mean is 48.38%. The total area between these scores is 95.52% (i.e., 44.14% + 48.38% = 95.52%). Thus, 95.52% of the area under the normal curve falls between scores of 85.0 and 111.0.

+2.14 and the mean (see Figure 5.7). To describe the total area between these two scores, then, we add 44.14% to 48.38% and learn that 92.52% of the area under curve falls between the raw scores of 85 and 111. As long as you understand the logic behind working with z scores and the z distribution, and you bother to draw a diagram similar to Figure 5.7 (for guidance, see Figure 5.3), answering the question is not at all difficult.

Area Between Two *z* Scores on the Same Side of the Distribution. Let's do a second example. This time, however, we will delineate the area between two scores that happen to fall on the same side of the mean. A teacher of gifted middle school students wants to know what percentage of IQ scores in the population fall between 132 and 138. As you may recall, standardized IQ tests usually have a μ of 100.0 and a σ of 15.0.

Right at the start, we know that these two scores are well into the right (upper) tail of the distribution. Before we begin to perform any calculations, a pressing issue of planning must be addressed: How can we determine the percentage distance between two scores that fall on the *same side of the distribution* (here, the positive side)? Previously, we have added relative distances together. Now, however, we must identify the area between the mean and where the two scores overlap one another—the remaining area under that section of the curve will represent the percentage of cases existing between the two IQ scores.

As always, begin by drawing a z distribution like the one shown in Figure 5.8 and then calculate the z scores corresponding to the two IQ scores. An IQ score of 132 is equal to a z of +2.13, or:

[5.10.1] $$z = \frac{132.0 - 100.0}{15} = \frac{32}{15} = +2.13.$$

The z score corresponding to an IQ score of 138 is:

[5.11.1] $$z = \frac{138.0 - 100.0}{15} = \frac{38}{15} = +2.53.$$

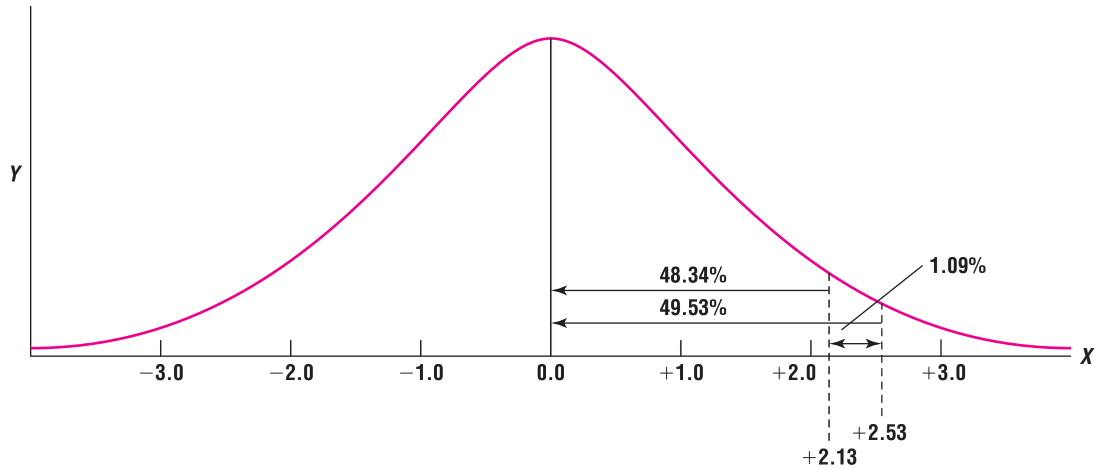


Figure 5.8 z Distribution Representing IQ Test with with $\mu = 100$ and $\sigma = 15$

Note: The raw scores of 132 and 138 are shown as z scores of $+2.13$ and $+2.53$, respectively. The area between $+2.13$ and the mean is 48.34%, and that between $+2.53$ and the mean is 49.43%. The total area between these scores is defined by subtracting the smaller area from the larger area (i.e., 49.43% − 48.34% = 1.09%). Thus, approximately 1% of the area under the curve—the possible cases—falls between scores of 132.0 and 138.0.

DATA BOX 5.C

Intelligence, Standardized IQ Scores, and the Normal Distribution

What is intelligence? How can it be measured and reduced to one score? What does the distribution of IQ scores look like?

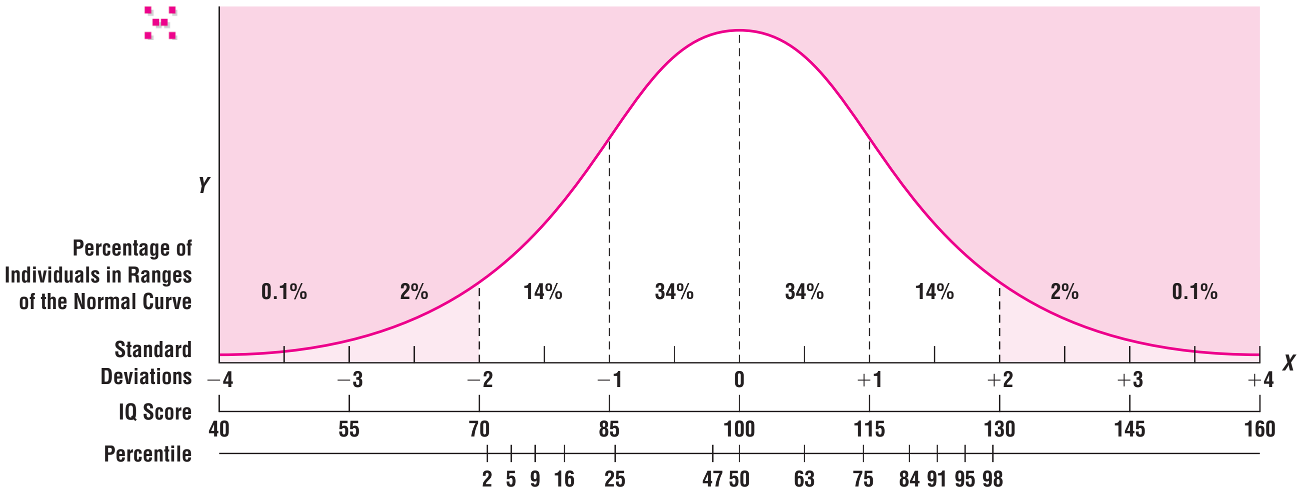
Defining the term "intelligence" is an old problem for educators and researchers, especially psychologists. Robert J. Sternberg, a prominent researcher who studies intelligence, defines it as the ability to learn from past experiences, to understand and control one's own thinking to promote learning, and to adapt to environments containing varied cultural and social elements (Sternberg, 1999; Sternberg & Detterman, 1986). To some extent, intelligence means different things to different people who find themselves (or the people they notice or, for psychologists, formally study) in varied contexts. A doctor diagnosing a medical problem can display a different type of diagnostic intelligence than a plumber repairing a drain or a mechanic fixing a plane engine, but each one demonstrates what can be called a contextual "intelligence."

Historically, psychology has been interested in assessing human intelligence via various intelligence tests, subsequently ranking people based on their scores. Although this practice has been subject to stark and accurate criticism (e.g., Gould, 1996), the use of intelligence testing and scores—notably the IQ score—continues. The original IQ score was devised by Stern (1912, cited in Sternberg, 1999), who argued that intelligence should be based on the ratio of an individual's mental age (MA) divided by chronological age (CA), multiplied by 100, or:

$$\text{IQ} = \frac{\text{MA}}{\text{CA}} \times 100.$$

If a teenager's mental age was the same as his chronological age (e.g., 13 years) then his IQ score would be 100—the exact average on the scale (i.e., IQ = [13/13] × 100 = 1.0 × 100 = 100). When mental age is greater than chronological age, the IQ score exceeds the average; when the reverse is true, the IQ score will fall below the mean of 100. This type of IQ score is called a *ratio IQ* (Sternberg, 1999).

Table 5.1 Normal Distribution of Deviation IQs



This figure shows a normal distribution as it applies to IQ, including identifying labels that are sometimes used to characterize different levels of IQ. It is important not to take these labels too seriously, as they are only loose characterizations, not scientific descriptions of performance.

Ratio IQs turned out to pose practical and conceptual measurement problems, so in recent decades they were replaced by *deviation IQ* scores. As you might guess, deviation IQ scores are based on the normal distribution and its assumptions for large populations. "Deviation" in this context means scores that deviate from the mean score within a normal distribution of IQ scores. The normal distribution of deviation IQs is shown in Table 5.1.

With any luck, you remain blissfully unaware of your measured IQ score and the above distribution is a mere curiosity for you. If you know your IQ score, however, then you are probably locating it on this curve to see how your performance compares to the population at large. In either case, you would do well to heed the sage wisdom of many research psychologists as well as intelligence test critics: The IQ score is but *one* measure of intelligence and an imperfect, incomplete one at that—no skill or behavior can be adequately reduced to one numerical index. As the definition provided earlier suggests, intelligence is comprised of various abilities. In other words, you are more than one score.

For further discussion of intelligence, its definition, and measurement, as well as accompanying controversies, see Ceci (1996), Gardner (1983), Gould (1996), and Sternberg (1985).

Locate and then plot where the two *z*s fall on your sketch of the *z* distribution (see Figure 5.8). Before proceeding any further, think about what relationship the plotted scores should disclose. First, both IQ scores are at least two standard deviation units away from the mean, placing both almost into the tail of the distribution. As a result, we should anticipate that the area between the two scores is probably quite small because few cases exist at the extreme of any distribution.

Turn to Table B.2 in Appendix B and locate the area between the mean and each of the two *z* scores. The percentages for +2.13 and +2.53, respectively, are 48.34% and 49.43%. As shown in Figure 5.8, then, a considerable amount of area is shared by the two scores (see the arrows pointing left from the *z* scores back to the mean)—only a small area exists between the scores (see the double-headed arrow pointing to both the scores). This small area represents the answer to the original question: What percentage of IQ scores on the test fall between 132 and 138?. To determine its numerical value, we subtract the smaller area associated with the z of $+2.13$ from the larger area for the z of $+2.53$, or $49.43\% - 48.34\% = 1.09\%$. In other words, only about 1% of the IQ scores in the distribution fall between scores of 132 and 138. Given that we know that relatively few people score in the upper ranges of the scale, this result makes sense. (As an aside, try to develop the practice of questioning any result—does it make statistical sense? Is it too large or too small?)

What if you needed to determine the area between two z scores that fell *below* the mean? The logic is the same as that used in the last example—you would simply be using negative z scores. To begin, of course, you would plot the standard scores, determine the area of overlap between each z and the mean, and then subtract the smaller area under the curve from the larger area. As an exercise, take a moment and make the sign on the z scores from the above example negative, sketch a z distribution, and then enter the scores on to it. Redo the above example using scores that now fall below the mean to verify that you understand the process and obtain the same answers.

A Further Transformed Score: The T Score

A variety of psychological and educational tests were created to adhere to the normal distribution and to yield z scores. Researchers, practitioners, and educators use such tests for academic as well as research purposes. Many times, use of actual z scores from these distributions is a problematic enterprise. Despite their benign qualities, some users can become confused by the presence of negative signs with some z scores. Moreover, these users will have forgotten—or perhaps never learned—the properties of z scores and the z distribution. Users can also be put off by the presence of numbers beyond the decimal point found even in many positive z scores, preferring to deal with whole numbers instead.

To deal with these and related concerns, tests originally based on transformations to z scores can be easily changed yet again to what are commonly called "transformed" or *T scores* (*not* to be confused with the t test, an inferential test reviewed extensively later in the book).

KEY TERM

A ***T* score** is a standard score which usually has a set μ and σ. Unlike z scores, *T scores* are reported as positive, whole numbers.

Transformed (T) scores are reported as positive, whole numbers.

A T score deals with the aforementioned decimal point problem by multiplying the z score by 10.0 or 100.0. The negative number problem, in turn, is eliminated by taking the resulting product and then adding it to or subtracting it from a positive whole number.

An example will demonstrate the efficacy of the T score concept. If we wished to work with a scale that had a set mean of 100.0, we could use the following T score formula from Runyon et al. (1996):

[5.12.1] $$T = \overline{T} + 10(z),$$

[5.12.2] $$T = 100 + 10(z).$$

where T is the transformed score, $\overline{T}$ is the mean of some scale determined by its creator or an investigator and $10(z)$ is a given z score multiplied by 10.

The formula shown above in [5.12.1] and [5.12.2] transforms any entered z scores into a new distribution with a mean of 100.0. Negative signs from negative zs are changed to positive values. Please note that this transformation simply presents the information in a different, easier to comprehend, way—the shape of the distribution and the relationships among the relative placement of scores on the test remain intact.

We can demonstrate the T score transformation by imagining an individual with a z score of -1.28 on some educational test. What is the corresponding T score? We simply enter the known z score into the T score formula, or:

[5.12.3] $$T = 100 + 10(-1.28),$$

[5.12.4] $$T = 100 - 12.8,$$

[5.12.5] $$T = 87.2.$$

Procedurally, the resulting *T* score is rounded to the nearest value of the whole number. Thus, a score of 87.2 would be reported as 87.0. Again, please note that the value of $\overline{T}$ will vary depending on the characteristics of the measure being used.

Can *T* scores be converted back to their *z* score equivalents? Yes, with little difficulty. The formula for this transformation back is:

[5.13.1] $$z = \frac{T - \overline{T}}{10.0}.$$

If we enter the original (nonrounded) *T* score, as well as $\overline{T}$, we find:

[5.13.2] $$z = \frac{87.2 - 100.0}{10.0}$$

[5.13.3] $$z = -\frac{12.8}{10.0}$$

[5.13.4] $$z = -1.28$$

Thus, we are back to the original *z* score.

Naturally, *z* scores derived from *T* scores are no different than any other *z* score. They, too, can be plotted on a *z* distribution or used to determine the percentile rank of a corresponding *T* score.

As noted by Runyon et al. (1996), the main contribution of *T* scores is that they permit researchers to compare the performance of test-takers from different subpopulations, such as disparate age groups. On a given test of motor skills, for instance, the test parameters of a college-aged population may differ from those associated with a much older group (age range 50 to 60 years). Once the *z*s for both groups are known, a *T* score transformation can convert them to the same distribution, allowing researchers to compare relative performance at one age with that found at another (e.g., Did younger individuals score higher on selected motor tasks relative to older persons?).

Writing About Standard Scores and the Normal Distribution

In a way, standard scores—whether *z* scores or *T* scores—are really the work horses, if you will, of descriptive statistics. Researchers and students are actually apt to *use* these standard scores in calculations more often than they are to *write* about them. When standardized measures are written up in a report or article, however, it is important to include either accompanying sample statistics ($\overline{X}, s$) or population parameters (μ, σ). Further, *z* and *T* scores must be appropriately used to answer, "compared to what?"

In a rare instance where you might be called on to write about *z* scores, for example, you would indicate the raw score and describe the test it was based on (i.e., statistics or parameters), and then indicate the score's use (e.g., to determine the area under the curve between the score and its mean, between two scores). If you were to describe a single score from a measure it might be structured as follows:

> The client received a raw score of 180 on the Youth Stress Index (YSI), a personality measure with a mean of 150 and a standard deviation of 15. Once converted to a standard score ($z = +2.00$), the client's YSI score was found to be two standard deviations above the mean, indicating a high level of adolescent stress. A score of 180 is at the 98th percentile of the YSI; only 2% of the adolescent population have been documented to have higher levels of reported stress at any point in time.

When writing about a z or T score, the goal is *not* to draw attention to the statistics per se, but to focus on the properties of the measure in question and the observations drawn from it.

Knowledge Base

1. You have a raw score of 220 drawn from a sample with a $\overline{X}$ of 215 and an s of 11.
   a. What is the z score corresponding to the raw score of 220?
   b. What proportion and percentage of cases occur between the mean and z?
   c. What proportion and percentage of cases exist beyond z?
2. What is the percentile rank of the raw score from question 1?
3. Using the sample statistics from question 1, determine the percentage of the area under the curve falling between scores of 210 and 218.
4. Using the sample statistics from question 1, determine the percentage of the area under the curve falling between scores of 200 and 213.
5. Using the z score transformation to a T score shown in formula [5.12.1] and [5.12.2], determine the T score for a z score of -2.20.

Answers

1. a. $+0.46$ b. .1772, 17.72% c. .3228, 32.28%
2. 67.72% of the cases fell at or below a score of 220.
3. 42.92%
4. 34.14%
5. T score $= 78.0$

Looking Ahead: Probability, *z* Scores, and the Normal Distribution

Probability is an attempt to predict the likelihood of particular events, usually those in the future. The study of probability is the formal examination of what event X is likely to happen given particular condition Y or Z. We will learn about probability in some detail in chapter 8. For the present, however, it is useful to know that z scores and the z distribution can be used to identify probabilities. We will introduce this topic in brief, conceptual terms here, supplying the numerical elements later.

Think about how the position of a given z score—its proximity to the mean, how many standard deviation units above or below the mean it falls—can define its typicality. In general, scores closer to the mean—the average—are more common, occur with greater frequency, than those further out in the tails (exceptions to this rule include bimodal distributions, among others). In a real sense, scores that lie closer to the mean are more likely to occur than any observations that fall in either tail. As we saw in an example from the previous section of the text, IQ scores falling between 132 and 138 are few and far between, comprising less than 2% of the scores in the entire population. Finding someone with an IQ of, say, 136 is a relative rarity; you are simply more likely to run into someone with an IQ of 102 because there are more people with IQs close to the distribution's average.

As standard scores that disclose relative placement within distributions, z scores are also used to discuss the typicality of observations. Think about it: If I randomly select a score from the z distribution, is it likely to be one that is within one, two, or three standard deviations of the mean in either direction? Given the bell shape of the normal distribution, I am relatively likely to draw an observation from the first standard deviation rather than the other two, just as I am more likely to select a score from the second rather than the third standard deviation from the mean (see Figure 5.4 to recall the percentages associated with each standard deviation under the normal curve).

Note the presence of the qualifying phrases "relatively likely" and "more likely" that get to the heart of the matter of probability. Probability is not certainty, so despite the fact that more scores cluster close to the mean, I might still randomly select one that lies two standard deviations or more away. In terms of chance, though, I am less likely to pick an extreme score. The relative likelihood underlying the selection of particular observations is the basis of probability theory. We will discuss how *z* scores can be used to determined specific numerical probabilities associated with the normal distribution in detail in chapter 8.

Project Exercise

UNDERSTANDING THE RECENTERING OF SCHOLASTIC APTITUDE TEST SCORES

There is a good chance that some readers of this textbook took the SAT I reasoning test prior to April 1995, when scores on the test were recentered. What does the term "recentering" mean? Essentially, recentering the SAT scores placed the average score on both the verbal and math subtests back to 500, the midpoint of the 200 to 800 scale. Why was this necessary? As noted in Data Box 5.B, the population of students taking the test has expanded greatly since 1941, the year that the scores of 10,000 college-bound students were used to establish norms—performance expectations and comparison scores—for the test.

Prior to this 1995 recentering of scores, the average SAT verbal had "fallen" from 500 to 424 and the math score changed from an average of 500 to 478. In other words, the distribution of test scores changed from 1941 to the 1990s. As shown graphically in Figure 5.9, the average score shifted from the 1941 midpoint to a point well below it. The distribution for the current (pre-1995) scale shown in Figure 5.9 appears to be showing some

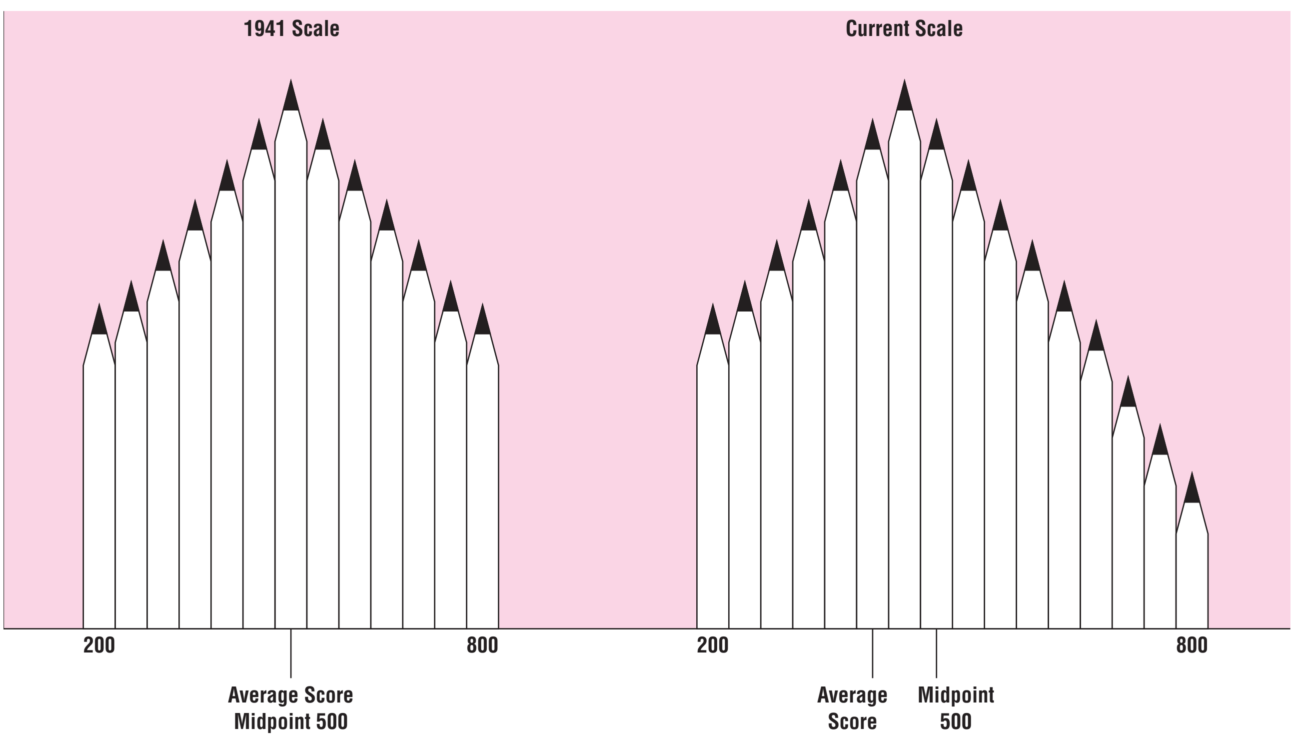


Figure 5.9 Shape of Distributions of Test Scores for 1941 and 1990s Student Samples

Note: The average score on the 1941 version of the verbal and math subscales of the SAT was 500. The average scores on these subscales in the early 1990s were 424 (verbal) and 478 (math), indicating that the population of test takers changed in the intervening years.

positive skew; that is, test scores appear to be clustering toward the lower end of the distribution. Recentering of the SAT made the distribution of scores once again appear like the normal (original) distribution shown in the left side of Figure 5.9.

By balancing scores on the SAT, the College Board returned the scores to a standard normal distribution. Following the logic of transforming scores presented in this chapter, "old" SAT scores were transformed into "new" ones. The difference here, of course, is that the shape of the distribution was changed somewhat—it was "normalized" once again, presumably by a transformation formula similar to one used for calculating *T* scores. Although the exact formula is not public information, I do have two equivalence tables that quickly illustrate how pre-April 1995 SAT scores on the verbal and math subtests can be changed to the "new" SAT scores (see Table 5.2).

Table 5.2 Conversion Tables From "Old" to "New" SAT I Scores

SAT verbal: original scale to recentered scale				SAT mathematical: original scale to recentered scale			
Equivalence Table				Equivalence Table			
Original Scale	Recentered Scale	Original Scale	Recentered Scale	Original Scale	Recentered Scale	Original Scale	Recentered Scale
800	800	500	580	800	800	500	520
790	800	490	570	790	800	490	520
780	800	480	560	780	800	480	510
770	800	470	550	770	790	470	500
760	800	460	540	760	770	460	490
750	800	450	530	750	760	450	480
740	800	440	520	740	740	440	480
730	800	430	510	730	730	430	470
720	790	420	500	720	720	420	460
710	780	410	490	710	700	410	450
700	760	400	480	700	690	400	440
690	750	390	470	690	680	390	430
680	740	380	460	680	670	380	430
670	730	370	450	670	660	370	420
660	720	360	440	660	650	360	410
650	710	350	430	650	650	350	400
640	700	340	420	640	640	340	390
630	690	330	410	630	630	330	380
620	680	320	400	620	620	320	370
610	670	310	390	610	610	310	350
600	670	300	380	600	600	300	340
590	660	290	370	590	600	290	330
580	650	280	360	580	590	280	310
570	640	270	350	570	580	270	300
560	630	260	340	560	570	260	280
550	620	250	330	550	560	250	260
540	610	240	310	540	560	240	240
530	600	230	300	530	550	230	220
520	600	220	290	520	540	220	200
510	590	210	270	510	530	210	200
		200	230			200	200

Source: The College Board, 45 Columbus Avenue, New York, NY 10023.

What happened to student scores? Generally, the conversion appeared to increase the scores of most students. Thus, someone who took the test in 1993 and scored a 510 verbal and a 560 math would now have scores of 590 and 570, respectively. On average, the College Board reports that most student scores increased as a result of the transformation. The rank and percentile rank of the scores remained virtually unchanged, however. Why? Recall what you know about relative placement in a distribution and conversion to standard scores—a given scores changes but the relative percent of test takes who score higher or lower remains the same.

Many admissions officers and educators feared that the recentering of scores meant that the SAT was suddenly an easier test to score well on. This fear was unfounded, of course, because the recentering merely placed the subtest means back to 500. Neither the inherent difficulty of the questions on the test nor the rank order of student scores on the test changed. Because everyone's score was readjusted, the relative placement of a given score compared to all other scores remained the same.

Here are some questions to think about and discuss regarding the transformation of scores and the normal distribution.

1. An admission's officer at a small university is worried that the recentering of the SAT will hurt student recruitment. Specifically, this professional is worried that potential applicants will be scared to apply because his institution's averages have increased—the scores printed in the university's admission's material seem to be much higher on the "new" SAT. As a student of statistics, what would you say to reduce his fears?
2. A college graduate happens to see the SAT conversion table, Table 5.2, converts her "old" verbal and math scores from 1992 to their "new" score equivalents. She then remarks to you that, "I always knew that they messed up my score on that test. I knew that I scored higher on that verbal test, especially—and look, I was right!" What do you tell this former student about the SAT transformation, as well as standard scores and the normal distribution more generally?
3. A peer examines the two distributions shown in Figure 5.9. Your peer comments that the two curves only confirm what he has always known—American students are getting dumber. "Just look at the clump of low scores in the curve on the right—and there are so few people who are at the upper end of the scale!," he says. What do you say in response? How can you cogently and concisely explain the statistical shift in scores from 1941 to the 1990s?

LOOKING FORWARD THEN BACK

Conceptually, standard or *z* scores are an integral part of statistical inference, helping to represent distinct variables with different qualities in a common frame of reference. The decision trees that open this chapter will help you to recall the role of *z* scores when working with distributions of data (decision tree one), converting raw scores into their *z* score equivalents (decision tree two), or using a *z* score to determine a percentile rank (decision tree three). Keep in mind that one of the *z* distribution's capabilities is the way it allows the data analyst to generalize. Knowing a score's placement enables analysts to make an inference about its relative relationship with other scores in a distribution. The decision trees will prove to be helpful whether you are in need of a quick refresher to recall the characteristics of *z* scores or to use them for descriptive purposes.

Summary

1. When evaluating any score or measure, it must always be compared to other observations—including the mean and standard deviation—within the distribution.
2. Standardizing scores or measures enables researchers to understand and compare them to other objects or observations along a single continuum.
3. Raw scores can be transformed into standard scores, which are an important part of various inferential statistical procedures. The apparent value of standard scores can change but no information, including their relation to other scores, is lost.
4. The *z* score is a descriptive statistic that identifies the distance between a raw score (X) and a distribution's mean, and it is reported in standard deviation units. A *z* score can be calculated as long as sample statistics ($\overline{X}$, s) or population parameters (μ, σ) are known. A given *z* score indicates how many standard deviations away from the mean the score is placed.
5. The *z* distribution has a mean of 0.0 and a standard deviation of 1.0, and it is symmetric; negative values fall below the mean and positive values lie above it. Conversion to *z* scores changes the value of scores but not their placement on or the shape of their parent distribution.
6. The standard normal distribution, a bell-shaped curve, contains observations laid out in a predictable pattern. The area under the curve (which can be known in proportion or percentage terms) is delineated in standard deviation units, such that the bulk of available observations are found—in descending order of magnitude—in the first, second, and third standard deviations to the right and left of the mean.
7. When allied with the standard normal distribution, *z* scores enable investigators to compare different types of measures, variables, or scores with one another on the same distribution. Such comparison is possible even when the means and standard deviations of the raw scores differ substantially from one another.
8. Due to their respective properties, *z* scores and the normal curve can be used to determine the percentile rank of a given raw score.
9. *T* scores are another form of standard score that usually converts *z* scores into positive, whole numbers.
10. Probability—predicting the likelihood of particular outcomes—is conceptually and numerically linked to *z* scores and the normal distribution.

Key Terms

Raw score *(p.182)*

Standard score *(p.182)*

T score *(p.196)*

z score *(p.183)*

Chapter Problems

1. What are the properties of the *z* distribution? Why are *z* scores useful?
2. Explain why comparison between one observation and another is so important for statistics and data analysis, as well as the behavioral sciences.
3. What are the properties of the standard normal distribution? Why is the normal curve useful for statisticians and behavioral scientists?
4. Explain what role the statistical concept of standard deviation plays in understanding *z* scores and the area under the normal curve.
5. Why do researchers standardize data? What is a standard score? Why are *z* scores and *T* scores standard scores?
6. If *z* scores appear to have such a wide range of applicability to statistical analysis, why did behavioral scientists find it necessary to develop *T* scores? What is the relationship between *z* and *T* scores?
7. What is probability? How does probability conceptually relate to *z* scores and the normal distribution?
8. A researcher examines her data and notices that its distribution is negatively skewed, and that its $\mu = 342.0$ and its $\sigma = 21.0$. If she converts all the data to *z* scores, what will be the numerical value of the mean and the standard deviation? What will the shape of the distribution of scores look like now?
9. Imagine you have a distribution with a $\mu = 78.0$ and its $\sigma = 12.5$. Find the *z* score equivalents of the following raw scores: 54.0, 63.5, 66.0, 77.0, 78.5, 81.0.
10. Imagine you have a distribution with a $\mu = 78.0$ and its $\sigma = 12.5$. Convert the following *z* scores back to their raw score equivalents: -3.10, -1.55, -1.0, $+0.55$, $+1.76$, $+2.33$, $+3.9$.
11. Sketch a *z* distribution and plot the *z* scores from problem 9. Determine the area under curve (in percent) between each *z* and the mean of the distribution.
12. Imagine you have a sample with a $\overline{X} = 35$ and an $s = 3.5$. Find the *z* score equivalents for the following raw scores: 27, 29.5, 34.3, 35, 45, 47.5.

13. Imagine you have a sample with a $\overline{X} = 35$ and an $s = 3.5$. Convert the following z scores back to their raw score equivalents: -3.01, -2.56, -1.21, $+1.40$, $+2.77$, 3.00.
14. Sketch a z distribution and plot the z scores from problem 12. Determine the area under curve (in percent) between each z and the mean of the distribution.
15. You have a very large distribution of IQ scores. As you know, the IQ test has a $\mu = 100.0$ and its $\sigma = 15.0$. Find the percentage of the curve falling between each of the following scores and the mean: 85, 88, 98.0, 112, 120, 133.
16. You have a very large distribution of IQ scores. As you know, the IQ test has a $\mu = 100.0$ and its $\sigma = 15.0$. Find the percentage of the curve falling between each of the following pairs of scores: 85 and 88; 92 and 96; 98 and 108; 115 and 128; 130 and 140; 141 and 143.
17. You have a data set with a $\overline{X} = 50.0$ and an $s = 7.0$. Find the percentage of the curve falling between each of the following scores and the mean: 38.0, 39.5, 45.0, 52.0, 57.0, 66.6.
18. You have a data set with a $\overline{X} = 50.0$ and an $s = 7.0$. Find the percentage of the curve falling between each of the following pairs of scores: 39.0 and 43.0; 44.5 and 56.5; 51.0 and 61.0; 62.0 and 63.0; 65.0 and 75.0; 76.0 and 80.0.
19. Determine what percentage of the cases under the normal curve are beyond each of the z scores calculated in problem 9.
20. Determine what percentage of the cases under the normal curve are beyond each of the z scores calculated in problem 12.
21. Determine what percentage of the cases under the normal curve are beyond each of the z scores calculated in problem 15.
22. Calculate the percentile rank of each of the z scores found in problem 9.
23. Calculate the percentile rank of each of the z scores found in problem 12.
24. Calculate the percentile rank of each of the z scores found in problem 15.
25. Are z scores normally distributed? Why or why not?
26. Is there one normal curve or many normal curves? Can there be more than one? How? Explain.
27. A student has taken four area tests designed to measure particular intellectual abilities. The following table identifies each test, summarizes its characteristics, and provides the student's score on it. Assume that the possible scores on each test are normally distributed.

Area Test	μ	σ	**Student's Score**
Verbal ability	58.5	6.50	63.0
Visualization	110.0	15.0	102.5
Memory	85.0	11.5	98.0
Spatial relations	320.0	33.5	343.0

   a. Change each of the student's score to its z score equivalent.
   b. On which test did the student receive a high score? A low score?
   c. What is the percentile rank of the student's verbal ability score? What percentage of the students who took the spatial relations test scored higher than the student?
28. Envision a normal distribution with $\mu = 88$ and $\sigma = 14$.
   a. Identify the scores at the 25th, 75th, and 95th percentiles.
   b. What percentage of cases fall below a score of 74.0?
   c. What percentage of scores are higher than a 93.0?
   d. What percentage of cases lie between the mean and a score of 99.0?
   e. What percentage of cases fall between scores of 86.5 and 92.5?
29. Using the T score formula of $T = 75 + 10(z)$, transform the following z scores into T scores. Round your answers to the nearest whole number: -1.12, -2.30, $+1.18$, $+2.67$, $+3.58$.
30. A transformation to T scores resulted in the following values when the formula $T = 400 + 100(z)$ was used. Convert these T scores back to their original z scores: 420, 373, 510, 485, 624.
31. Use the decision trees opening this chapter to answer the following questions:
   a. An investigator wants to know a data point's location relative to the mean of its (large) distribution. What should the investigator do?
   b. A researcher draws an observation from an intact population whose parameters are known. How can this observation be converted to a z score?
   c. What is the percentile rank of a z score equal to 1.76?