2-6 Measures of Relative Standing

This section introduces measures that can be used to compare values from different data sets, or to compare values within the same data set. We introduce z scores (for comparing values from different data sets) and quartiles and percentiles (for comparing values within the same data set).

z Scores

A z score (or standard score) is found by converting a value to a standardized scale, as given in the following definition. We will use z scores extensively in Chapter 5 and later chapters, so they are extremely important.

Definition

A **standard score,** or ***z* score,** is the number of standard deviations that a given value x is above or below the mean. It is found using the following expressions:

Sample

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

or

Population

$$z = \frac{x - \mu}{\sigma}$$

(Round z to two decimal places.)

The following example illustrates how z scores can be used to compare values, even though they might come from different populations.

EXAMPLE Comparing Heights NBA superstar Michael Jordan is 78 in. tall and WNBA basketball player Rebecca Lobo is 76 in. tall. Jordan is obviously taller by 2 in., but which player is *relatively* taller? Does Jordan's height among men exceed Lobo's height among women? Men have heights with a mean of 69.0 in. and a standard deviation of 2.8 in.; women have heights with a mean of 63.6 in. and a standard deviation of 2.5 in. (based on data from the National Health Survey).

SOLUTION To compare the heights of Michael Jordan and Rebecca Lobo relative to the populations of men and women, we need to standardize those heights by converting them to z scores.

$$\text{Jordan:} \quad z = \frac{x - \mu}{\sigma} = \frac{78 - 69.0}{2.8} = 3.21$$

$$\text{Lobo:} \quad z = \frac{x - \mu}{\sigma} = \frac{76 - 63.6}{2.5} = 4.96$$

INTERPRETATION Michael Jordan's height is 3.21 standard deviations above the mean, but Rebecca Lobo's height is a whopping 4.96 standard deviations above the mean. Rebecca Lobo's height among women is relatively greater than Michael Jordan's height among men.

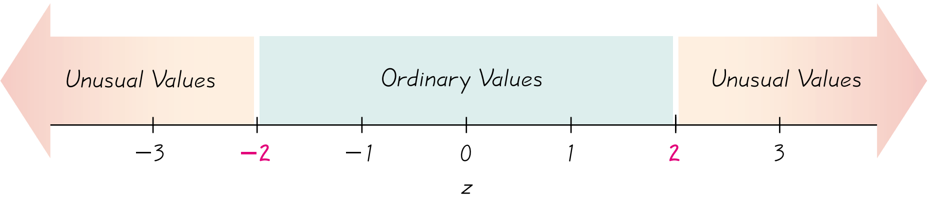
z Scores and Unusual Values

In Section 2-5 we used the range rule of thumb to conclude that a value is "unusual" if it is more than 2 standard deviations away from the mean. It follows that unusual values have z scores less than -2 or greater than $+2$. (See Figure 2-14 on page 94.) Using this criterion, both Michael Jordan and Rebecca Lobo are unusually tall because they both have heights with z scores greater than 2.

While considering professional basketball players with exceptional heights, another player is Mugsy Bogues, who was successful even though he is only 5 ft 3 in. tall. (We again use the fact that men have heights with a mean of 69.0 in. and a

FIGURE 2-14 Interpreting *z* scores

Unusual values are those with *z* scores less than 2.00 or greater than 2.00.



standard deviation of 2.8 in.) After converting 5 ft 3 in. to 63 in., we convert his height to a *z* score as follows:

$$\text{Bogues: } z = \frac{x - \mu}{\sigma} = \frac{63 - 69.0}{2.8} = -2.14$$

Let's be grateful to Mugsy Bogues for his many years of inspired play and for illustrating this principle:

Whenever a value is less than the mean, its corresponding *z* score is negative.

Ordinary values: $-2 \le z \text{ score} \le 2$

Unusual values: $z \text{ score} < -2$ *or* $z \text{ score} > 2$

z scores are measures of position in the sense that they describe the location of a value (in terms of standard deviations) relative to the mean. A *z* score of 2 indicates that a value is two standard deviations *above* the mean, and a *z* score of −3 indicates that a value is three standard deviations *below* the mean. Quartiles and percentiles are also measures of position, but they are defined differently than *z* scores and they are useful for comparing values within the same data set or between different sets of data.

Quartiles and Percentiles

Recall from Section 2-4 that the median of a data set is the middle value, so that 50% of the values are equal to or less than the median and 50% of the values are greater than or equal to the median. Just as the median divides the data into two equal parts, the three **quartiles,** denoted by Q_1, Q_2, and Q_3, divide the sorted values into four equal parts. (Values are *sorted* when they are arranged in order.) Here are descriptions of the three quartiles:

Q_1 (First quartile): Separates the bottom 25% of the sorted values from the top 75% (To be more precise, at least 25% of the sorted values are less than or equal to Q_1, and at least 75% of the values are greater than or equal to Q_1.)

Q_2 (Second quartile): Same as the median; separates the bottom 50% of the sorted values from the top 50%.

Q_3 (Third quartile): Separates the bottom 75% of the sorted values from the top 25%. (To be more precise, at least 75% of the sorted values are less than or equal to Q_3, and at least 25% of the values are greater than or equal to Q_3.)

We will describe a procedure for finding quartiles after we discuss percentiles. There is not universal agreement on a single procedure for calculating quartiles, and different computer programs often yield different results. For example, if you use the data set of 1, 3, 6, 10, 15, 21, 28, and 36, you will get these results:

	Q_1	Q_2	Q_3
STATDISK	4.5	12.5	24.5
Minitab	3.75	12.5	26.25
Excel	5.25	12.5	22.75
TI-83 Plus	4.5	12.5	24.5

For this data set, STATDISK and the TI-83 Plus calculator agree, but they do not always agree. If you use a calculator or computer software for exercises involving quartiles, you may get results that differ slightly from the answers given in the back of the book.

Just as there are three quartiles separating a data set into four parts, there are also 99 **percentiles,** denoted $P_1, P_2, \ldots, P_{99}$, which partition the data into 100 groups with about 1% of the values in each group. (Quartiles and percentiles are examples of *quantiles*—or *fractiles*—which partition data into groups with roughly the same number of values.)

The process of finding the percentile that corresponds to a particular value x is fairly simple, as indicated in the following expression:

$$\text{percentile of value } x = \frac{\text{number of values less than } x}{\text{total number of values}} \cdot 100$$

EXAMPLE Cotinine Levels of Smokers Table 2-13 lists the 40 sorted cotinine levels of smokers included in Table 2-1. Find the percentile corresponding to the cotinine level of 112.

SOLUTION From Table 2-13 we see that there are 12 values less than 112, so

$$\text{percentile of } 112 = \frac{12}{40} \cdot 100 = 30$$

INTERPRETATION The cotinine level of 112 is the 30th percentile.

The preceding example shows how to convert from a given sample value to the corresponding percentile. There are several different methods for the reverse procedure of converting a given percentile to the corresponding value in the data set. The procedure we will use is summarized in Figure 2-15, which uses the notation that follows the figure.

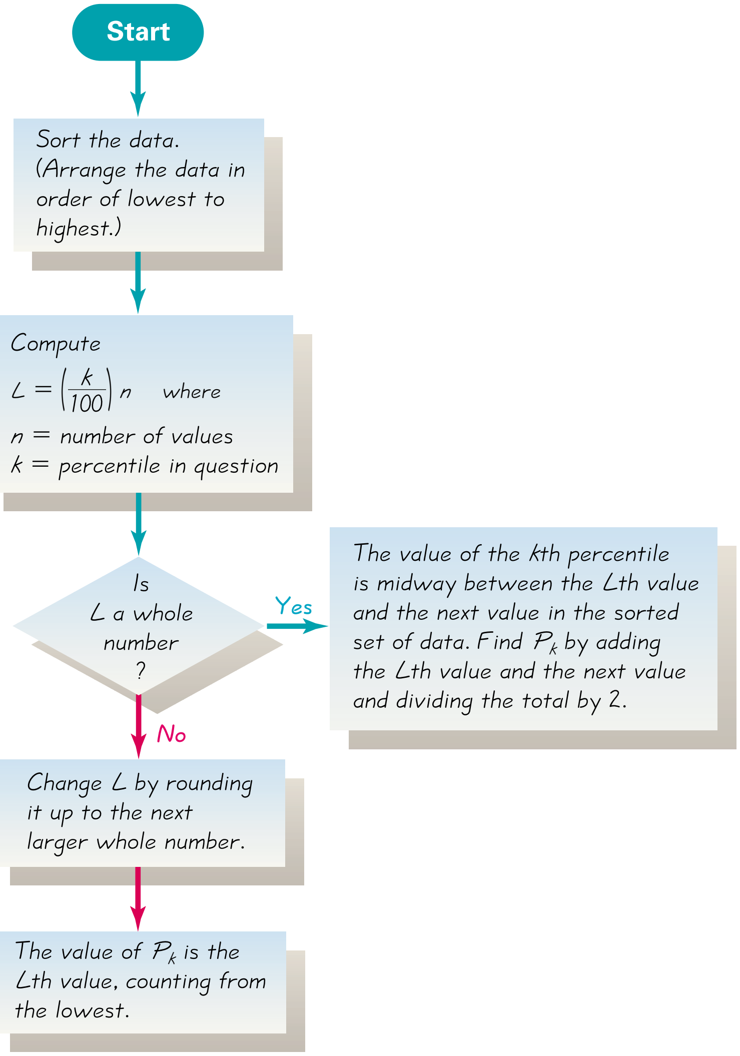
Table 2-13 *Sorted* Cotinine Levels of 40 Smokers

0	1	1	3	17	32	35	44	48	86
87	103	112	121	123	130	131	149	164	167
173	173	198	208	210	222	227	234	245	250
253	265	266	277	284	289	290	313	477	491

Cost of Laughing Index

There really is a Cost of Laughing Index (CLI), which tracks costs of such items as rubber chickens, Groucho Marx glasses, admission to comedy clubs, and 13 other leading humor indicators. This is the same basic approach used in developing the Consumer Price Index (CPI), which is based on a weighted average of goods and services purchased by typical consumers. While standard scores and percentiles allow us to compare different values, they ignore any element of time. Index numbers, such as the CLI and CPI, allow us to compare the value of some variable to its value at some base time period. The value of an index number is the current value, divided by the base value, multiplied by 100.

FIGURE 2-15 Converting from the *k*th Percentile to the Corresponding Data Value



Notation

n = total number of values in the data set
k = percentile being used (Example: For the 25th percentile, $k = 25$.)
L = locator that gives the *position* of a value (Example: For the 12th value in the sorted list, $L = 12$.)
P_k = kth percentile (Example: P_{25} is the 25th percentile.)

EXAMPLE Cotinine Levels of Smokers Refer to the sorted cotinine levels of smokers in Table 2-13 and use Figure 2-15 to find the value of the 68th percentile, P_{68}.

SOLUTION Referring to Figure 2-15, we see that the sample data are already sorted, so we can proceed to find the value of the locator L. In this computation we use $k = 68$ because we are trying to find the value of the 68th percentile. We use $n = 40$ because there are 40 data values.

$$L = \frac{k}{100} \cdot n = \frac{68}{100} \cdot 40 = 27.2$$

Next, we are asked if L is a whole number and we answer no, so we proceed to the next lower box where we change L by rounding it up from 27.2 to 28. (In this book we typically round off the usual way, but this is one of two cases where we round *up* instead of rounding *off*.) Finally, the bottom box shows that the value of P_{68} is the 28th value, counting up from the lowest. In Table 2-13, the 28th value is 234. That is, $P_{68} = 234$.

EXAMPLE Cotinine Levels of Smokers Refer to the sample of cotinine levels of smokers given in Table 2-13. Use Figure 2-15 to find the value of Q_1, which is the first quartile.

SOLUTION First we note that Q_1 is the same as P_{25}, so we can proceed with the objective of finding the value of the 25th percentile. Referring to Figure 2-15, we see that the sample data are already sorted, so we can proceed to compute the value of the locator L. In this computation, we use $k = 25$ because we are attempting to find the value of the 25th percentile, and we use $n = 40$ because there are 40 data values.

$$L = \frac{k}{100} \cdot n = \frac{25}{100} \cdot 40 = 10$$

Next, we are asked if L is a whole number and we answer yes, so we proceed to the box located at the right. We now see that the value of the kth (25th) percentile is midway between the Lth (10th) value and the next value in the original set of data. That is, the value of the 25th percentile is midway between the 10th value and the 11th value. The 10th value is 86 and the 11th value is 87, so the value midway between them is 86.5. We conclude that the 25th percentile is $P_{25} = 86.5$. The value of the first quartile Q_1 is also 86.5.

The preceding example showed that when finding a quartile value (such as Q_1), we can use the equivalent percentile value (such as P_{25}) instead. See the margin for relationships relating quartiles to equivalent percentiles.

$Q_1 = P_{25}$
$Q_2 = P_{50}$
$Q_3 = P_{75}$

In earlier sections of this chapter we described several statistics, including the mean, median, mode, range, and standard deviation. Some other statistics are defined using quartiles and percentiles, as in the following:

$$\text{interquartile range (or IQR)} = Q_3 - Q_1$$

$$\text{semi-interquartile range} = \frac{Q_3 - Q_1}{2}$$

$$\text{midquartile} = \frac{Q_3 + Q_1}{2}$$

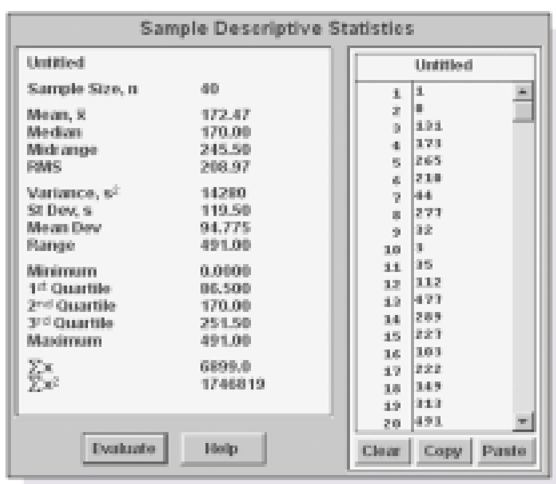
$$\text{10–90 percentile range} = P_{90} - P_{10}$$

After completing this section, you should be able to convert a value into its corresponding z score (or standard score) so that you can compare it to other values, which may be from different data sets. You should be able to convert a value into its corresponding percentile value so that you can compare it to other values in some data set. You should be able to convert a percentile to the corresponding data value. And finally, you should understand the meanings of quartiles and be able to relate them to their corresponding percentile values (as in $Q_3 = P_{75}$).

Using Technology

A variety of different computer programs and calculators can be used to find many of the statistics introduced so far in this chapter. In Section 2-4 we provided specific instructions for using STATDISK, Minitab, Excel, and the TI-83 Plus calculator. We noted that we can sometimes enter a data set and use one operation to get several different sample statistics, often referred to as *descriptive statistics*. Examples of such results are shown in the following screen displays. These screen displays result from using the cotinine levels of the *smokers* given in Table 2-1 with the Chapter Problem. The TI-83 Plus results are shown on two screens because they do not all fit on one screen.

STATDISK



Minitab

Variable	N	Mean	Median	TrMean	StDev	SE Mean
SMOKER	40	172.5	170.0	164.7	119.5	18.9

Variable	Minimum	Maximum	Q1	Q3
SMOKER	0.0	491.0	86.3	252.3

Excel

Column1	
Mean	172.475
Standard Error	18.89434
Median	170
Mode	1
Standard Deviation	119.4983
Sample Variance	14279.85
Kurtosis	0.519621
Skewness	0.587929
Range	491
Minimum	0
Maximum	491
Sum	6899
Count	40

TI-83 Plus

```
1-Var Stats
 x̄=172.475
 Σx=6899
 Σx²=1746819
 Sx=119.4983076
 σx=117.9951244
↓n=40
```

```
1-Var Stats
↑n=40
 minX=0
 Q1=86.5
 Med=170
 Q3=251.5
 maxX=491
```

2-6 Basic Skills and Concepts

In Exercises 1–4, express all z scores with two decimal places.

1. IQ Scores Stanford Binet IQ scores have a mean of 100 and a standard deviation of 16. Albert Einstein reportedly had an IQ of 160.
a. What is the difference between Einstein's IQ and the mean?
b. How many standard deviations is that [the difference found in part (a)]?
c. Convert Einstein's IQ score to a z score.
d. If we consider "usual" IQ scores to be those that convert to z scores between -2 and 2, is Einstein's IQ usual or unusual?

2. Pulse Rates of Adults Assume that adults have pulse rates (beats per minute) with a mean of 72.9 and a standard deviation of 12.3 (based on data from the National Health Examination). When this exercise question was written, the author's pulse rate was 48.
a. What is the difference between the author's pulse rate and the mean?
b. How many standard deviations is that [the difference found in part (a)]?
c. Convert a pulse rate of 48 to a z score.
d. If we consider "usual" pulse rates to be those that convert to z scores between -2 and 2, is a pulse rate of 48 usual or unusual? Can you explain why a pulse rate might be unusually low? (The reason for this low pulse rate is *not* that statistics textbook authors are usually in a state that could loosely be described as comatose.)

3. Heights of Men Adult males have heights with a mean of 69.0 in. and a standard deviation of 2.8 in. Find the z scores corresponding to the following:
a. Actor Danny DeVito, who is 5 ft tall
b. NBA basketball player Shaquille O'Neal, who is 7 ft 1 in. tall
c. The author, who is a 69.72-in.-tall golf and tennis "player"

4. Body Temperatures Human body temperatures have a mean of 98.20° and a standard deviation of 0.62°. Convert the given temperatures to z scores.
a. 100° **b.** 96.96° **c.** 98.20°

In Exercises 5–8, express all z scores with two decimal places. Consider a score to be unusual if its z score is less than −2.00 or greater than 2.00.

5. Heights of Women The Beanstalk Club is limited to women and men who are very tall. The minimum height requirement for women is 70 in. Women's heights have a mean of 63.6 in. and a standard deviation of 2.5 in. Find the z score corresponding to a woman with a height of 70 in. and determine whether that height is unusual.

6. Length of Pregnancy A woman wrote to *Dear Abby* and claimed that she gave birth 308 days after a visit from her husband, who was in the Navy. Lengths of pregnancies have a mean of 268 days and a standard deviation of 15 days. Find the z score for 308 days. Is such a length unusual? What do you conclude?

7. Body Temperature Human body temperatures have a mean of 98.20° and a standard deviation of 0.62°. An emergency room patient is found to have a temperature of 101°. Convert 101° to a z score. Is that temperature unusually high? What does it suggest?

8. Cholesterol Levels For men aged between 18 and 24 years, serum cholesterol levels (in mg/100 ml) have a mean of 178.1 and a standard deviation of 40.7 (based on data from the National Health Survey). Find the z score corresponding to a male, aged 18–24 years, who has a serum cholesterol level of 259.0 mg/100 ml. Is this level unusually high?

9. Comparing Test Scores Which is relatively better: A score of 85 on a psychology test or a score of 45 on an economics test? Scores on the psychology test have a mean of 90 and a standard deviation of 10. Scores on the economics test have a mean of 55 and a standard deviation of 5.

10. Comparing Scores Three students take equivalent tests of a sense of humor and, after the laughter dies down, their scores are calculated. Which is the highest relative score?
a. A score of 144 on a test with a mean of 128 and a standard deviation of 34.
b. A score of 90 on a test with a mean of 86 and a standard deviation of 18.
c. A score of 18 on a test with a mean of 15 and a standard deviation of 5.

T 11. Weights of Coke Refer to Data Set 17 in Appendix B for the sample of 36 weights of regular Coke. Convert the weight of 0.7901 to a z score. Is 0.7901 an unusual weight for regular Coke?

T 12. Green M&Ms Refer to Data Set 19 in Appendix B for the sample of weights of green M&M candies. Convert the weight of the heaviest green M&M candy to a z score. Is the weight of that heaviest green M&M an unusual weight for green M&Ms?

In Exercises 13–16, use the 40 sorted cotinine levels of smokers listed in Table 2-13. Find the percentile corresponding to the given cotinine level.

13. 149 **14.** 210 **15.** 35 **16.** 250

In Exercises 17–24, use the 40 sorted cotinine levels of smokers listed in Table 2-13. Find the indicated percentile or quartile.

17. P_{20} **18.** Q_3 **19.** P_{75} **20.** Q_2
21. P_{33} **22.** P_{21} **23.** P_1 **24.** P_{85}

In Exercises 25–28, use the cholesterol levels of females listed in Data Set 1 of Appendix B. Find the percentile corresponding to the given cholesterol level.

25. 123 **26.** 309 **27.** 271 **28.** 126

In Exercises 29–36, use the cholesterol levels of females listed in Data Set 1 of Appendix B. Find the indicated percentile or quartile.

29. P_{85} **30.** P_{35} **31.** Q_1 **32.** Q_3

33. P_{18} **34.** P_{36} **35.** P_{58} **36.** P_{96}

2-6 Beyond the Basics

37. Units of Measurement When finding a z score for the height of a basketball player in the NBA, how is the result affected if, instead of using inches, all heights are expressed in centimeters? In general, how are z scores affected by the particular unit of measurement that is used?

38. Converting a z Score Heights of women have a mean of 63.6 in. and a standard deviation of 2.5 in.

a. Julia Roberts, who is one of the most successful actresses in recent years, has a height that converts to a z score of 2.16. How tall (in inches) is Julia Roberts?

b. Female rapper Lil' Kim has a height that converts to a z score of -1.84. How tall (in inches) is Lil' Kim?

39. Distribution of z Scores

a. A data set has a distribution that is uniform. If all of the values are converted to z scores, what is the shape of the distribution of the z scores?

b. A data set has a distribution that is bell-shaped. If all of the values are converted to z scores, what is the shape of the distribution of the z scores?

c. In general, how is the shape of a distribution affected if all values are converted to z scores?

40. Fibonacci Sequence Here are the first several terms of the famous Fibonacci sequence: 1, 1, 2, 3, 5, 8, 13.

a. Find the mean $\bar{x}$ and standard deviation s, then convert each value to a z score. Don't round the z scores; carry as many places as your calculator can handle.

b. Find the mean and standard deviation of the z scores found in part (a).

c. If you use any other data set, will you get the same results obtained in part (b)?

41. Cotinine Levels of Smokers Use the sorted cotinine levels of smokers listed in Table 2-13.

a. Find the interquartile range.

b. Find the midquartile.

c. Find the 10–90 percentile range.

d. Does $P_{50} = Q_2$? If so, does P_{50} *always* equal Q_2?

e. Does $Q_2 = (Q_1 + Q_3)/2$? If so, does Q_2 *always* equal $(Q_1 + Q_3)/2$?

42. Interpolation When finding percentiles using Figure 2-15, if the locator L is not a whole number, we round it up to the next larger whole number. An alternative to this procedure is to interpolate so that a locator of 23.75 leads to a value that is 0.75 (or 3/4) of the way between the 23rd and 24th values. Use this method of interpolation to find P_{35} and Q_1 for the weights of bears listed in Data Set 9 of Appendix B.

43. Deciles and Quintiles For a given data set, there are nine **deciles,** denoted by D_1, $D_2, \ldots, D_9$ which separate the sorted data into 10 groups, with about 10% of the values in each group. There are also four **quintiles,** which divide the sorted data into five groups, with about 20% of the values in each group. (Note the difference between quintiles and quantiles, which were described earlier in this section.)

a. Which percentile is equivalent to D_1? D_5? D_8?

b. Using the sorted cotinine levels of smokers in Table 2-13, find the nine deciles.

c. Using the sorted cotinine levels of smokers in Table 2-13, find the four quintiles.