

2.6 INTERPRETING THE STANDARD DEVIATION

We've seen that if we are comparing the variability of two samples selected from a population, the sample with the larger standard deviation is the more variable of the two. Thus, we know how to interpret the standard deviation on a relative or comparative basis, but we haven't explained how it provides a measure of variability for a single sample.

To understand how the standard deviation provides a measure of variability of a data set, consider a specific data set and answer the following questions: How many measurements are within 1 standard deviation of the mean? How many measurements are within 2 standard deviations? For example, look at the 100 mileage per gallon readings given in Table 2.3. Recall that $\bar{x} = 36.99$ and $s = 2.42$. Then

$$\begin{aligned}\bar{x} - s &= 34.57 & \bar{x} + s &= 39.41 \\ \bar{x} - 2s &= 32.15 & \bar{x} + 2s &= 41.83\end{aligned}$$

If we examine the data, we find that 68 of the 100 measurements, or 68%, are in the interval

$$\bar{x} - s \text{ to } \bar{x} + s$$

Similarly, we find that 96, or 96%, of the 100 measurements are in the interval

$$\bar{x} - 2s \text{ to } \bar{x} + 2s$$

We usually write these intervals as

$$(\bar{x} - s, \bar{x} + s) \text{ and } (\bar{x} - 2s, \bar{x} + 2s)$$

Such observations identify criteria for interpreting a standard deviation that apply to *any* set of data, whether a population or a sample. The criteria, expressed as a mathematical theorem and as a rule of thumb, are presented in Tables 2.7 and 2.8. In these tables we give two sets of answers to the questions of how many measurements fall within 1, 2, and 3 standard deviations of the mean. The first, which applies to *any* set of data, is derived from a theorem proved by the Russian mathematician P. L. Chebyshev (1821–1894). The second, which applies to symmetric, **mound-shaped distributions** of data (where the mean, median, and mode are all about the same), is based upon empirical evidence that has accumulated over the years. However, the percentages given for the intervals in Table 2.8 provide remarkably good approximations even when the distribution of the data is slightly skewed or asymmetric.

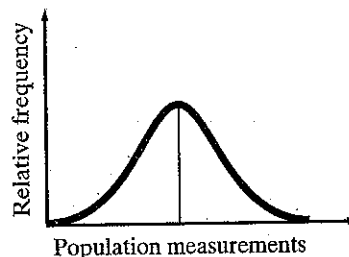
TABLE 2.7 Interpreting the Standard Deviation: Chebyshev's Rule

Chebyshev's Rule applies to any data set, regardless of the shape of the frequency distribution of the data.

- No useful information is provided on the fraction of measurements that fall within 1 standard deviation of the mean, i.e., within the interval $(\bar{x} - s, \bar{x} + s)$ for samples and $(\mu - \sigma, \mu + \sigma)$ for populations.
- At least $\frac{3}{4}$ of the measurements will fall within 2 standard deviations of the mean, i.e., within the interval $(\bar{x} - 2s, \bar{x} + 2s)$ for samples and $(\mu - 2\sigma, \mu + 2\sigma)$ for populations.
- At least $\frac{8}{9}$ of the measurements will fall within 3 standard deviations of the mean, i.e., within the interval $(\bar{x} - 3s, \bar{x} + 3s)$ for samples and $(\mu - 3\sigma, \mu + 3\sigma)$ for populations.
- Generally, for any number k greater than 1, at least $(1 - 1/k^2)$ of the measurements will fall within k standard deviations of the mean, i.e., within the interval $(\bar{x} - ks, \bar{x} + ks)$ for samples and $(\mu - k\sigma, \mu + k\sigma)$ for populations.

TABLE 2.8 Interpreting the Standard Deviation: The Empirical Rule

The **Empirical Rule** is a rule of thumb that applies to data sets with frequency distributions that are mound-shaped and symmetric, as shown below.



- Approximately 68% of the measurements will fall within 1 standard deviation of the mean, i.e., within the interval $(\bar{x} - s, \bar{x} + s)$ for samples and $(\mu - \sigma, \mu + \sigma)$ for populations.
- Approximately 95% of the measurements will fall within 2 standard deviations of the mean, i.e., within the interval $(\bar{x} - 2s, \bar{x} + 2s)$ for samples and $(\mu - 2\sigma, \mu + 2\sigma)$ for populations.
- Approximately 99.7% (essentially all) of the measurements will fall within 3 standard deviations of the mean, i.e., within the interval $(\bar{x} - 3s, \bar{x} + 3s)$ for samples and $(\mu - 3\sigma, \mu + 3\sigma)$ for populations.

EXAMPLE 2.10

Thirty students in an experimental psychology class use various techniques to train a rat to move through a maze. At the end of the course, each student's rat is timed through the maze. The results (in minutes) are listed in Table 2.9. Determine the fraction of the 30 measurements in the intervals $\bar{x} \pm s$, $\bar{x} \pm 2s$, and $\bar{x} \pm 3s$, and compare the results with those predicted in Tables 2.7 and 2.8.

● **RATMAZE**

TABLE 2.9 Times (in Minutes) of 30 Rats Running Through a Maze

1.97	.60	4.02	3.20	1.15	6.06	4.44	2.02	3.37	3.65
1.74	2.75	3.81	9.70	8.29	5.63	5.21	4.55	7.60	3.16
3.77	5.36	1.06	1.71	2.47	4.25	1.93	5.15	2.06	1.65

Solution

First, we entered the data into the computer and used STATISTIX to produce summary statistics. The mean and standard deviation of the sample data, highlighted on the printout shown in Figure 2.19, are (rounded)

$$\bar{x} = 3.74 \text{ minutes} \quad s = 2.20 \text{ minutes}$$

FIGURE 2.19

STATISTIX Printout for Example 2.10

DESCRIPTIVE STATISTICS	
	RUNTIME
N	30
MEAN	3.7443
SD	2.1982
VARIANCE	4.8323
MINIMUM	0.6000
MEDIAN	3.5100
MAXIMUM	9.7000

Now, we form the interval

$$(\bar{x} - s, \bar{x} + s) = (3.74 - 2.20, 3.74 + 2.20) = (1.54, 5.94)$$

A check of the measurements shows that 23 of the times are within this 1 standard deviation interval around the mean. This number represents $23/30$, or $\approx 77\%$ of the sample measurements.

The next interval of interest is

$$(\bar{x} - 2s, \bar{x} + 2s) = (3.74 - 4.40, 3.74 + 4.40) = (-.66, 8.14)$$

All but two of the times are within this interval, so $28/30$, or approximately 93%, are within 2 standard deviations of \bar{x} .

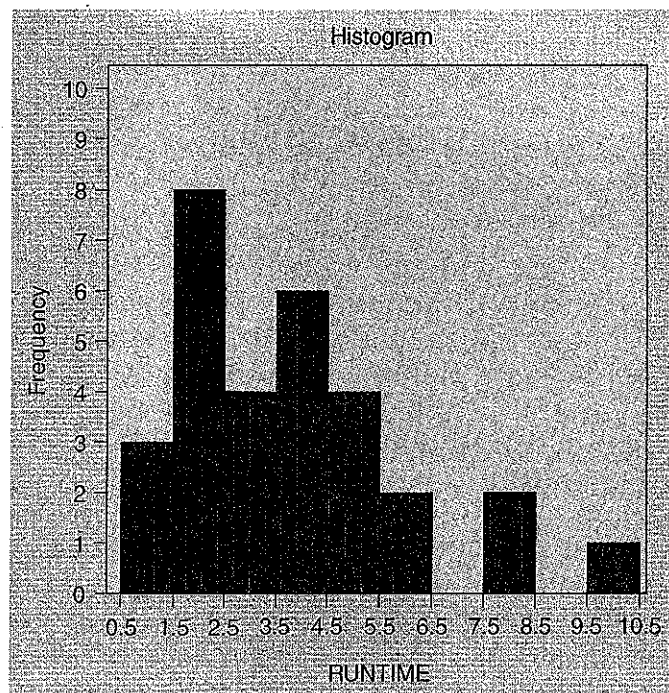
Finally, the 3-standard-deviation interval around \bar{x} is

$$(\bar{x} - 3s, \bar{x} + 3s) = (3.74 - 6.60, 3.74 + 6.60) = (-2.86, 10.34)$$

All of the times fall within 3 standard deviations of the mean.

These 1-, 2-, and 3-standard-deviation percentages (77%, 93%, and 100%) agree fairly well with the approximations of 68%, 95%, and 100% given by the Empirical Rule (Table 2.8) for mound-shaped distributions. If you look at the STATISTIX frequency histogram for this data set in Figure 2.20, you'll note that the distribution is not really mound-shaped, nor is it extremely skewed. Thus, we get reasonably good results from the mound-shaped approximations. Of course, we know from Chebyshev's Rule (Table 2.7) that no matter what the shape of the distribution, we would expect at least 75% and 89% of the measurements to lie within 2 and 3 standard deviations of \bar{x} , respectively.

FIGURE 2.20
STATISTIX Histogram for
Times for Rats to Move Through
Maze



EXAMPLE 2.11

Chebyshev's Rule and the Empirical Rule (Tables 2.7 and 2.8) are useful as a check on the calculation of the standard deviation. For example, suppose we calculated the standard deviation for the gas mileage data (Table 2.3) to be 5.85. Are there any "clues" in the data that enable us to judge whether this number is reasonable?

Solution

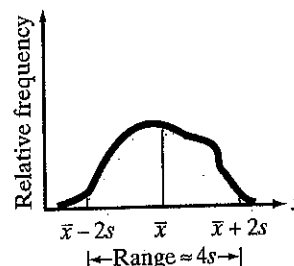
The range of the mileage data in Table 2.3 is $44.0 - 30.0 = 14.9$. From Chebyshev's Rule and the Empirical Rule we know that most of the measurements (approximately 95% if the distribution is mound-shaped) will be within 2 standard deviations of the mean. And, regardless of the shape of the distribution and the number of measurements, almost all of them will fall within 3 standard deviations of the mean. Consequently, we would expect the range of the measurements to be between 4 (i.e., $\pm 2s$) and 6 (i.e., $\pm 3s$) standard deviations in length (see Figure 2.21). For the car mileage data, this means that s should fall between

$$\frac{\text{Range}}{6} = \frac{14.9}{6} = 2.48 \quad \text{and} \quad \frac{\text{Range}}{4} = \frac{14.9}{4} = 3.73$$

In particular, the standard deviation should not be much larger than $1/4$ of the range, particularly for the data set with 100 measurements. Thus, we have reason to believe that the calculation of 5.85 is too large. A check of our work reveals that 5.85 is the variance s^2 , not the standard deviation s (see Example 2.9). We "forgot" to take the square root (a common error); the correct value is $s = 2.42$. Note that this value is slightly smaller than the range divided by 6 (2.48). The larger the data set, the greater the tendency for very large or very small measurements (extreme values) to appear, and when they do, the range may exceed 6 standard deviations. ■

FIGURE 2.21

The Relation Between the Range and the Standard Deviation



In examples and exercises we'll sometimes use $s \approx \text{range}/4$ to obtain a crude, and usually conservatively large, approximation for s . However, we stress that this is no substitute for calculating the exact value of s when possible.

Finally, and most importantly, we will use the concepts in Chebyshev's Rule and the Empirical Rule to build the foundation for statistical inference-making. The method is illustrated in Example 2.12.

EXAMPLE 2.12

A manufacturer of automobile batteries claims that the average length of life for its grade A battery is 60 months. However, the guarantee on this brand is for just 36 months. Suppose the standard deviation of the life length is known to be 10 months, and the frequency distribution of the life-length data is known to be mound-shaped.

- Approximately what percentage of the manufacturer's grade A batteries will last more than 50 months, assuming the manufacturer's claim is true?

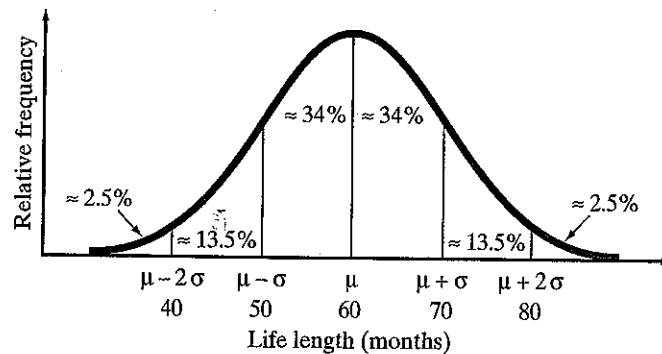
- b. Approximately what percentage of the manufacturer's batteries will last less than 40 months, assuming the manufacturer's claim is true?
- c. Suppose your battery lasts 37 months. What could you infer about the manufacturer's claim?

Solution

If the distribution of life length is assumed to be mound-shaped with a mean of 60 months and a standard deviation of 10 months, it would appear as shown in Figure 2.22. Note that we can take advantage of the fact that mound-shaped distributions are (approximately) symmetric about the mean, so that the percentages given by the Empirical Rule can be split equally between the halves of the distribution on each side of the mean. The approximations given in Figure 2.22 are more dependent on the assumption of a mound-shaped distribution than those given by the Empirical Rule (Table 2.8), because the approximations in Figure 2.22 depend on the (approximate) symmetry of the mound-shaped distribution. We saw in Example 2.10 that the Empirical Rule can yield good approximations even for skewed distributions. This will *not* be true of the approximations in Figure 2.22; the distribution *must* be mound-shaped and (approximately) symmetric.

FIGURE 2.22

Battery Life-Length
Distribution: Manufacturer's
Claim Assumed True



For example, since approximately 68% of the measurements will fall within 1 standard deviation of the mean, the distribution's symmetry implies that approximately $(1/2)(68\%) = 34\%$ of the measurements will fall between the mean and 1 standard deviation on each side. This concept is illustrated in Figure 2.22. The figure also shows that 2.5% of the measurements lie beyond 2 standard deviations in each direction from the mean. This result follows from the fact that if approximately 95% of the measurements fall within 2 standard deviations of the mean, then about 5% fall outside 2 standard deviations; if the distribution is approximately symmetric, then about 2.5% of the measurements fall beyond 2 standard deviations on each side of the mean.

- a. It is easy to see in Figure 2.22 that the percentage of batteries lasting more than 50 months is approximately 34% (between 50 and 60 months) plus 50% (greater than 60 months). Thus, approximately 84% of the batteries should have life length exceeding 50 months.
- b. The percentage of batteries that last less than 40 months can also be easily determined from Figure 2.22. Approximately 2.5% of the batteries should fail prior to 40 months, assuming the manufacturer's claim is true.
- c. If you are so unfortunate that your grade A battery fails at 37 months, you can make one of two inferences: Either your battery was one of the approximately 2.5% that fail prior to 40 months, or something about the manufacturer's claim is not true. Because the chances are so small that a battery fails before 40 months, you would have good reason to have serious doubts about

the manufacturer's claim. A mean smaller than 60 months and/or a standard deviation longer than 10 months would both increase the likelihood of failure prior to 40 months.*

Example 2.12 is our initial demonstration of the statistical inference-making process. At this point you should realize that we'll use sample information (in Example 2.12, your battery's failure at 37 months) to make inferences about the population (in Example 2.12, the manufacturer's claim about the life length for the population of all batteries). We'll build on this foundation as we proceed.



EXERCISES 2.66-2.81

Learning the Mechanics

- 2.66** To what kind of data sets can Chebyshev's Rule be applied? The Empirical Rule?
- 2.67** The output from a statistical computer program indicates that the mean and standard deviation of a data set consisting of 200 measurements are \$1,500 and \$300, respectively.
- What are the units of measurement of the variable of interest? Based on the units, what type of data is this: quantitative or qualitative?
 - What can be said about the number of measurements between \$900 and \$2,100? Between \$600 and \$2,400? Between \$1,200 and \$1,800? Between \$1,500 and \$2,100?
- 2.68** For any set of data, what can be said about the percentage of the measurements contained in each of the following intervals?
- $\bar{x} - s$ to $\bar{x} + s$
 - $\bar{x} - 2s$ to $\bar{x} + 2s$
 - $\bar{x} - 3s$ to $\bar{x} + 3s$
- 2.69** For a set of data with a mound-shaped relative frequency distribution, what can be said about the percentage of the measurements contained in each of the intervals specified in Exercise 2.68?
- 2.70** The following is a sample of 25 measurements:
- | | | | | | | | | | | | | |
|---|----|---|----|---|---|----|----|----|----|---|---|---|
| 7 | 6 | 6 | 11 | 8 | 9 | 11 | 9 | 10 | 8 | 7 | 7 | 5 |
| 9 | 10 | 7 | 7 | 7 | 7 | 9 | 12 | 10 | 10 | 8 | 6 | |
- Compute \bar{x} , s^2 , and s for this sample.
 - Count the number of measurements in the intervals $\bar{x} \pm s$, $\bar{x} \pm 2s$, and $\bar{x} \pm 3s$. Express each count as a percentage of the total number of measurements.

- Compare the percentages found in part **b** to the percentages given by the Empirical Rule and Chebyshev's Rule.
 - Calculate the range and use it to obtain a rough approximation for s . Does the result compare favorably with the actual value for s found in part **a**?
- 2.71** Given a data set with a largest value of 760 and a smallest value of 135, what would you estimate the standard deviation to be? Explain the logic behind the procedure you used to estimate the standard deviation. Suppose the standard deviation is reported to be 25. Is this feasible? Explain.

Applying the Concepts—Basic

- 2.72** To minimize the potential for gastrointestinal disease outbreaks, all passenger cruise ships arriving at U.S. ports are subject to unannounced sanitation inspections. Ships are rated on a 100-point scale by the Centers for Disease Control and Prevention. A score of 86 or higher indicates that the ship is providing an accepted standard of sanitation. The May 2001 sanitation scores for 151 cruise ships are listed in the accompanying table, followed by a MINITAB printout of descriptive statistics on p. 68.
- Locate the mean and standard deviation on the MINITAB printout.
 - Calculate the intervals $\bar{x} \pm s$, $\bar{x} \pm 2s$, $\bar{x} \pm 3s$.
 - Find the percentage of measurements in the data set that fall within each of the intervals, part **b**. Do these percentages agree with either Chebyshev's Theorem or the Empirical Rule?

*The assumption that the distribution is mound-shaped and symmetric may also be incorrect. However, if the distribution were skewed to the right, as life-length distributions often tend to be, the percentage of measurements more than 2 standard deviations below the mean would be even less than 2.5%.