Comparing Variation in Different Populations

In Part 1 of this section, we noted that when comparing variation in two different sets of data, the standard deviations should be compared only if the two sets of data use the same scale and units and have approximately the same mean. If the means are substantially different, or if the samples use different scales or measurement units, we can use the *coefficient of variation,* defined as follows.

DEFINITION The **coefficient of variation** (or **CV**) for a set of nonnegative sample or population data, expressed as a percent, describes the standard deviation relative to the mean, and is given by the following:

Sample	Population
$CV = \frac{s}{\bar{x}} \cdot 100\%$	$CV = \frac{\sigma}{\mu} \cdot 100\%$

Round-Off Rule for the Coefficient of Variation

Round the coefficient of variation to one decimal place (such as 18.3%).

Example 8 Chocolate Chip Cookies and Coke

Compare the variation in the numbers of chocolate chips in Chips Ahoy (regular) cookies (listed in Table 3-1) and the weights of regular Coke listed in Data Set 19 of Appendix B. Using the sample data, we have these results: Cookies have $\bar{x} = 24.0$ chips and $s = 2.6$ chips; Coke has $\bar{x} = 0.81682$ lb and $s = 0.00751$ lb. Note that we want to compare variation among *numbers of chocolate chips* to variation among *weights of Coke*.

Solution

We can compare the standard deviations if the same scales and units are used and the two means are approximately equal, but here we have different scales (numbers of chocolate chips and weights of Coke) and different units of measurement (numbers and pounds), so we use the coefficients of variation:

Numbers of chocolate chips: $$CV = \frac{s}{\bar{x}} \cdot 100\% = \frac{2.6 \text{ chocolate chips}}{24.0 \text{ chocolate chips}} \cdot 100\% = 10.8\%$$

Weights of Coke: $$CV = \frac{s}{\bar{x}} \cdot 100\% = \frac{0.00751 \text{ lb}}{0.81682 \text{ lb}} \cdot 100\% = 0.9\%$$

Although the standard deviation of 2.6 chocolate chips cannot be compared to the standard deviation of 0.00751 lb, we can compare the coefficients of variation, which have no units. We can see that the numbers of chocolate chips (with $CV = 10.8\%$) vary considerably more than weights of Coke (with $CV = 0.9\%$). This makes intuitive sense, because variation among the numbers of chocolate chips is not a big deal, but if some cans of Coke were underfilled or overfilled by large amounts, the result would be angry consumers.

Biased and Unbiased Estimators

The sample standard deviation s is a **biased estimator** of the population standard deviation σ. This means that values of the sample standard deviation s do *not* target