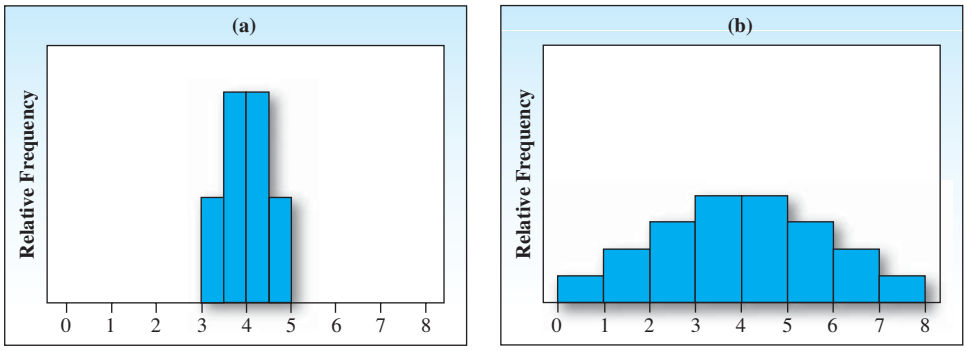
2.3 MEASURES OF VARIABILITY

Data sets may have the same center but look different because of the way the numbers *spread out* from the center. Consider the two distributions shown in Figure 2.6. Both distributions are centered at $x = 4$, but there is a big difference in the way the measurements spread out, or *vary*. The measurements in Figure 2.6(a) vary from 3 to 5; in Figure 2.6(b) the measurements vary from 0 to 8.

FIGURE 2.6
Variability or dispersion of data



Variability or **dispersion** is a very important characteristic of data. For example, if you were manufacturing bolts, extreme variation in the bolt diameters would cause a high percentage of defective products. On the other hand, if you were trying to discriminate between good and poor accountants, you would have trouble if the examination always produced test grades with little variation, making discrimination very difficult.

Measures of variability can help you create a mental picture of the spread of the data. We will present some of the more important ones. The simplest measure of variation is the **range.**

Definition The **range, *R,*** of a set of n measurements is defined as the difference between the largest and smallest measurements.

For the birth weight data in Table 1.9, the measurements vary from 5.6 to 9.4. Hence, the range is $9.4 - 5.6 = 3.8$. The range is easy to calculate, easy to interpret, and is an adequate measure of variation for small sets of data. But, for large data sets, the range is not an adequate measure of variability. For example, the two relative frequency distributions in Figure 2.7 have the same range but very different shapes and variability.

FIGURE 2.7
Distributions with equal range and unequal variability