0-9 Measures of Center, Spread, and Position

Objective: Find measures of center, spread, and position.

Statistics is the science of collecting, organizing, displaying, and analyzing data in order to draw conclusions and make predictions.

The entire group of interest to a statistician is called a **population**. A **variable** is a characteristic of a population that can assume different values called **data**. Data that include only one variable are called **univariable data**. When it is not possible to obtain data about every member of a population, a representative **sample** or subset of the population is selected.

Univariable data are often summarized using a single number to represent what is average or typical. Measure of average are also called **measures of center** or **central tendency**. The most common measures of center are mean, median, and mode.

KeyConcept Measures of Center

- The **mean** is the sum of the values in a set of data $x_1, x_2, \ldots, x_n$ divided by the total number of values n in the set. The formula for population mean μ is $\mu = \frac{x_1 + x_2 + \ldots + x_n}{n}$.
- The **median** is the middle value or the mean of the two middle values in a set of data when the data are arranged in numerical order.
- The **mode** is the value or values that appear most often in a set of data. A set of data can have no mode, one mode, or more than one mode.

Example 1

The number of milligrams of sodium in a 12-ounce can of ten different brands of regular cola are shown below. Find the mean, median, and mode.

50, 30, 25, 20, 40, 35, 35, 10, 15, 35

Because two very different data sets can have the same mean, statisticians also use **measures of spread** or **variation** to describe how widely the data values vary and how much the values differ from what is typical.

KeyConcept Measures of Spread

- The **range** is the difference between the greatest and least values in a set of data.
- The **variance** in a set of data $x_1, x_2, \ldots, x_n$ is the mean of the squares of the deviations or differences from the mean. The formula for population variance σ^2 is

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_n - \mu)^2}{n}.$$

- The **standard deviation** in a set of data $x_1, x_2, \ldots, x_n$ is the average amount by which each individual value deviates or differs from the mean. It is the square root of the variance. The formula for population standard deviation σ is

$$\sigma = \sqrt{\sigma^2} \text{ or } \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_n - \mu)^2}{n}}.$$

Example 2

Two classes took the same midterm exam. The scores of five students from each class are shown. Both sets of scores have a mean of 84.2.

Class A
85, 76, 92, 88, 80

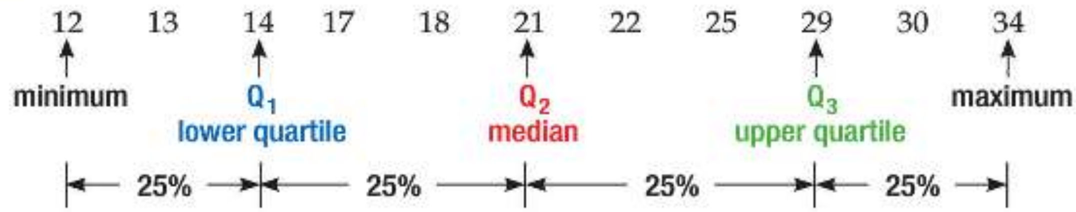
Class B
75, 85, 95, 98, 68

a) Find the range, variance, and standard deviation from the sample scores from Class A.

b) Find the range, variance, and standard deviation from the sample scores from Class B.

c) Compare the sample standard deviations of Class A and Class B.

Statisticians use measures of position to describe where specific values fall within a data set. **Quartiles** are three position measures that divide a data set arranged in ascending order into four groups, each containing about 25% of the data. The median marks the second quartile Q_2 and separates the data into upper and lower halves.
The first or **lower quartile** Q_1 is the median of the lower half, while the third or **upper quartile** Q_3 is the median of the upper half.



The three quartiles, along with the minimum and maximum values, are called a **five-number summary** of a data set.

<u>Example 3</u>

The number of hours Liana worked each week for the last 12 weeks were 21, 10, 18, 12, 15, 13, 20, 20, 19, 16, 18, and 14.

a) Find the minimum, lower quartile, median, upper quartile, and maximum of the data set.

b) Interpret this five-number summary.

The difference between Q_3 and Q_1, is called the **interquartile range** IQR. The IQR contains about 50% of the values. Before deciding on which measures best describe a set of data, check the data set for outliers. An **outlier** is an extremely high or extremely low value when compared with the rest of the values in the set. To check for outliers, look for data values that are beyond the upper or lower quartiles by more than 1.5 times the IQR.

Example 4

The number of minutes each of the 22 students in a class spent working on the same algebra assignment is shown below.
15, 12, 25, 15, 27, 10, 16, 18, 30, 35, 22, 25, 65, 20, 18, 25, 15, 13, 25, 22, 15, 28

a) Identify any outliers in the data.
b) Find the mean, median, mode, range, and standard deviation of the data set with and without the outlier. Describe the effect on each measure.