Unit 7, Lesson 5

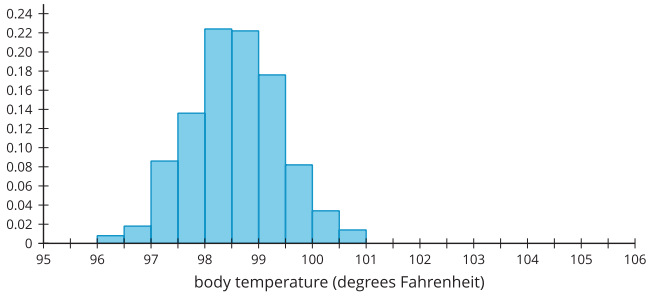
# Normal Distributions

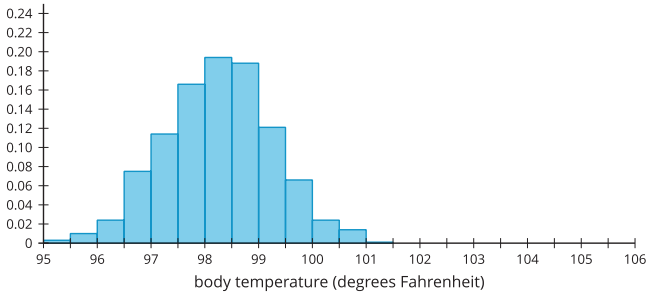
Let's investigate a specific type of distribution called a "normal distribution."

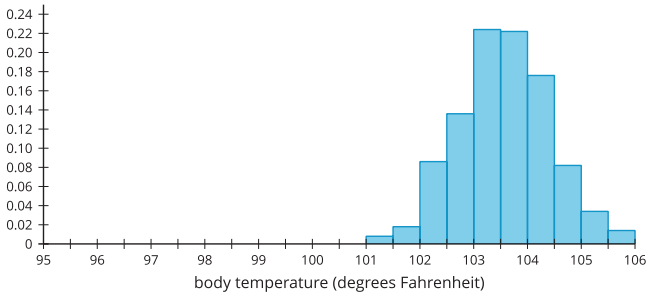
## 5.1 Body Temperature

Each histogram represents a group of 500 healthy people who had their temperature taken. Three histograms represent examples of data that approximate a **normal distribution**, and three histograms represent data that do not approximate a normal distribution. What do you think are the elements that define a normal distribution?

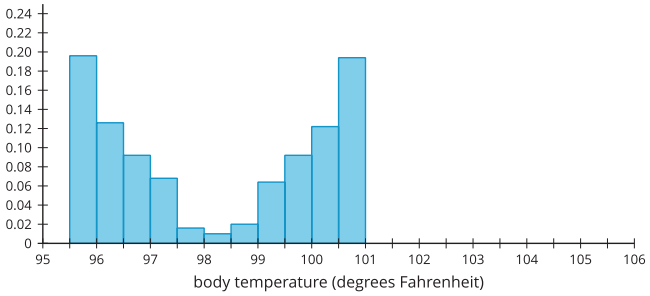
Some approximately normal distributions:

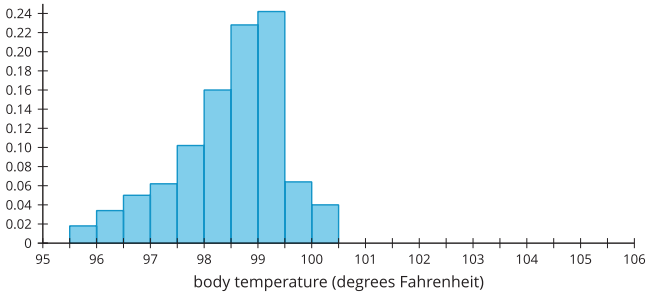


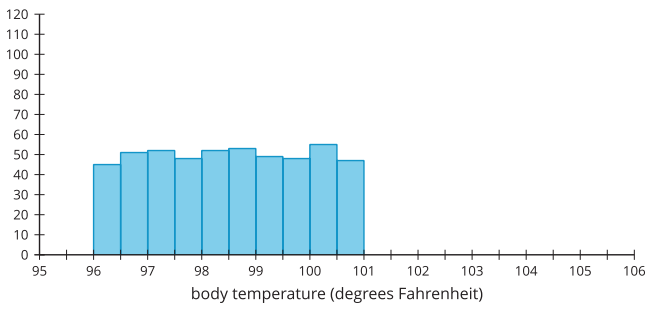



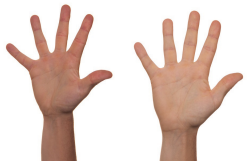

Some distributions that are not approximately normal distributions:



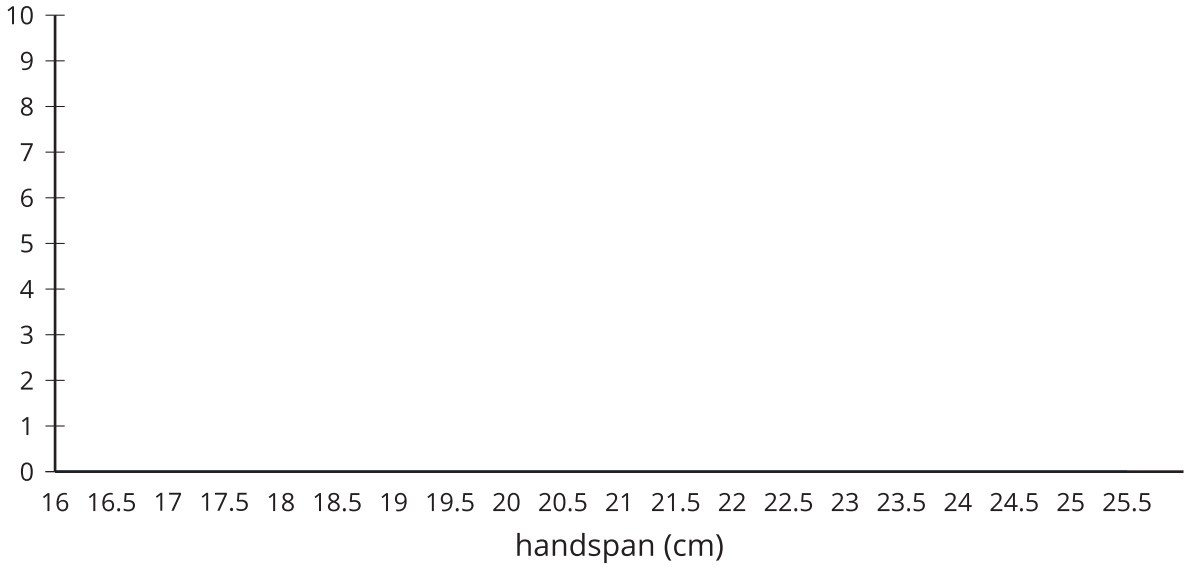
# 5.2 Playing a Piano



On some pianos, the average distance from one white key to the next is 2.39 centimeters. How many of your classmates could reach two notes that are 9 keys apart (21.5 cm) on a keyboard using only one hand?

1. Stretch your fingers apart as wide as you can, and measure the farthest distance from your thumb to your smallest finger. Round your measurement to the nearest tenth of a centimeter.

2. Your teacher will collect the measurements from the class. Draw a dot plot or histogram from the class data.




3. Describe the distribution you drew using terms such as “symmetric,” “approximately symmetric,” “skewed left,” “skewed right,” “approximately uniform,” “uniform,” “bell-shaped,” or “bimodal.” Estimate the center of your distribution.

4. How would you use your distribution to determine how many people in the class can reach the two notes 9 keys apart?

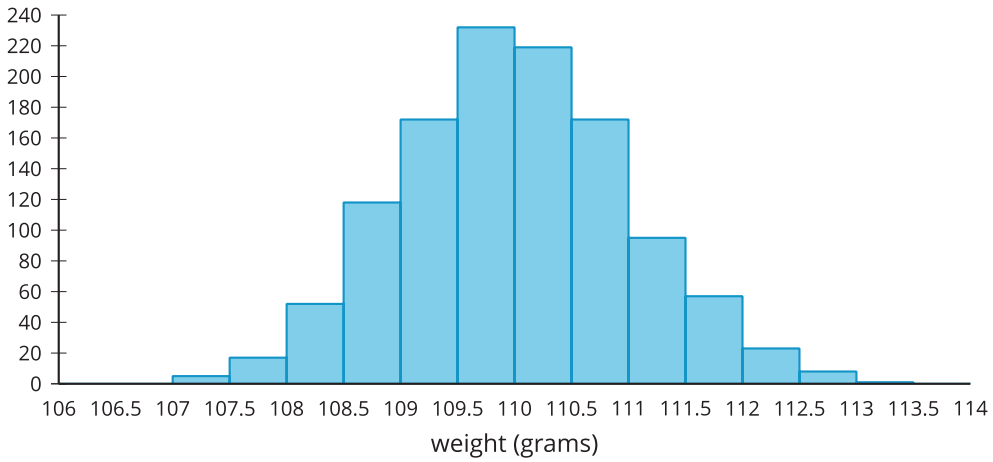
# 5.3 Relative Frequency Distribution

Manufacturers of butter make sticks of butter that weigh 110 grams on average. A manufacturer suspects the machine that forms the sticks of butter may have a problem, so they weigh each stick of butter the machine produces in an hour.

The weights are grouped into intervals of 0.5 gram and are summarized in a frequency table.

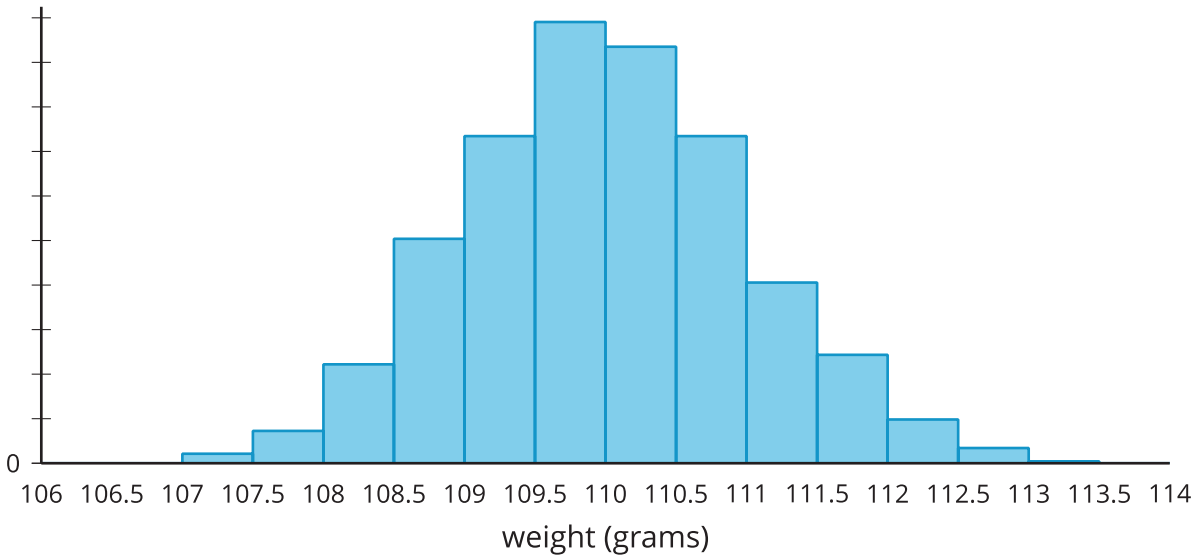
| weight (grams) | frequency | relative frequency |
|---|---|---|
| 107–107.5 | 5 | 0.0043 |
| 107.5–108 | 17 | |
| 108–108.5 | 52 | |
| 108.5–109 | 118 | |
| 109–109.5 | 172 | |
| 109.5–110 | 232 | |
| 110–110.5 | 219 | |
| 110.5–111 | 172 | |
| 111–111.5 | 95 | |
| 111.5–112 | 57 | |
| 112–112.5 | 23 | |
| 112.5–113 | 8 | |
| 113–113.5 | 1 | |
| total | 1,171 | |

The same data are summarized in this histogram.



Although this information is useful, it might be more helpful to know the proportion of sticks of butter in each weight interval rather than the actual number of sticks in that weight interval.

1. Complete the table by dividing each frequency value by the total number of sticks of butter in the data set. Round each value to 4 decimal places.
2. A **relative frequency histogram** is a histogram in which the height of each bar is the relative frequency. Since the heights of the bars are found by dividing each height by the total number of sticks of butter, the shape of the distribution is the same as a regular histogram, but the labels on the $y$-axis are changed. Label the $y$-axis with the correct values for each mark.

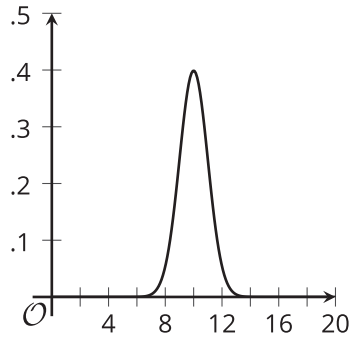



3. The manufacturer believes they should replace parts of the machine if more than 25% of the sticks of butter are more than 1 gram away from the intended value of 110 grams.
   a. Indicate on the relative frequency histogram the bars that correspond to sticks of butter that are more than 1 gram away from the intended weight of a stick of butter.
   b. Should this machine be replaced? Explain or show your reasoning.

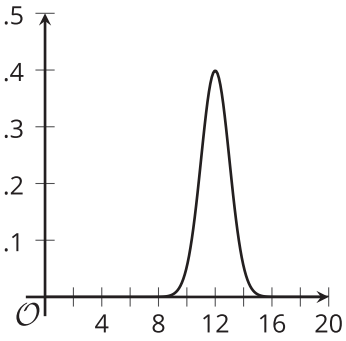
# 5.4 The Normal Curve

These curves represent normal distributions with different means and standard deviations. What do you notice?

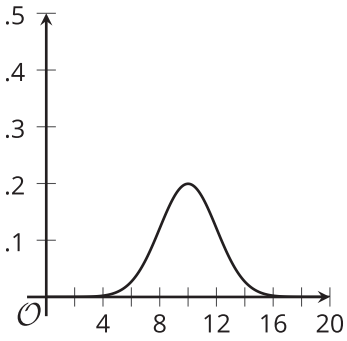
**mean: 10, standard deviation: 1**



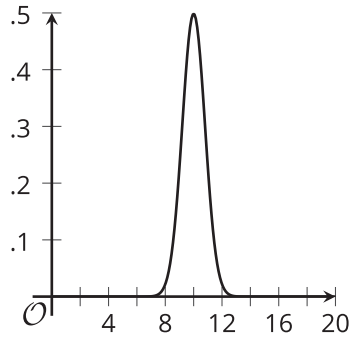

**mean: 12, standard deviation: 1**



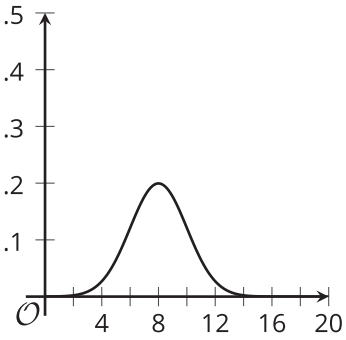

**mean: 10, standard deviation: 2**




**mean: 10, standard deviation: 0.8**




**mean: 8, standard deviation: 2**




## Are you ready for more?

The equation for the curve for a normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where $\sigma$ is the standard deviation, $\mu$ is the mean, and $e$ is a particular number close to 2.718.

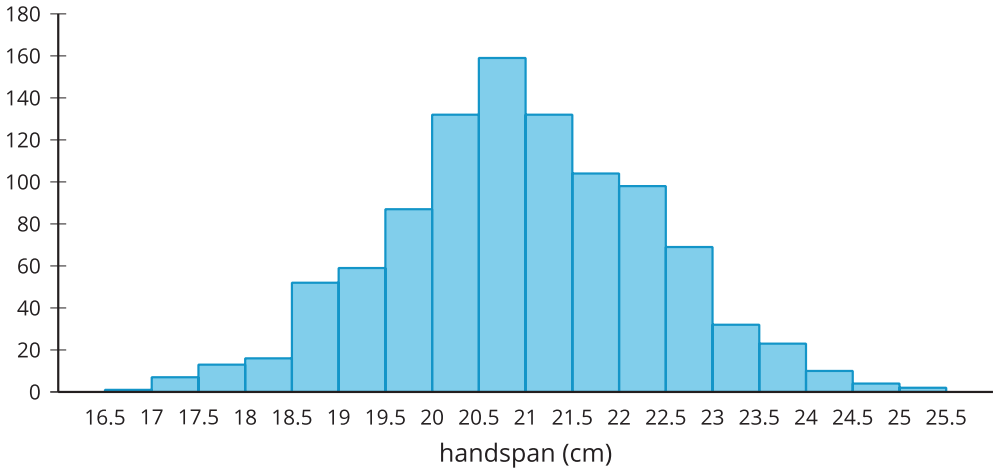
Notice the part of the function that is $(x - \mu)$. Using your understanding of transformations of functions, how does changing $\mu$ affect the graph of the function?

## Lesson 5 Summary

A histogram shows the number of items in a data set that fall into specified intervals. A **relative frequency histogram** shows the proportion of an entire data set that falls into specified intervals.

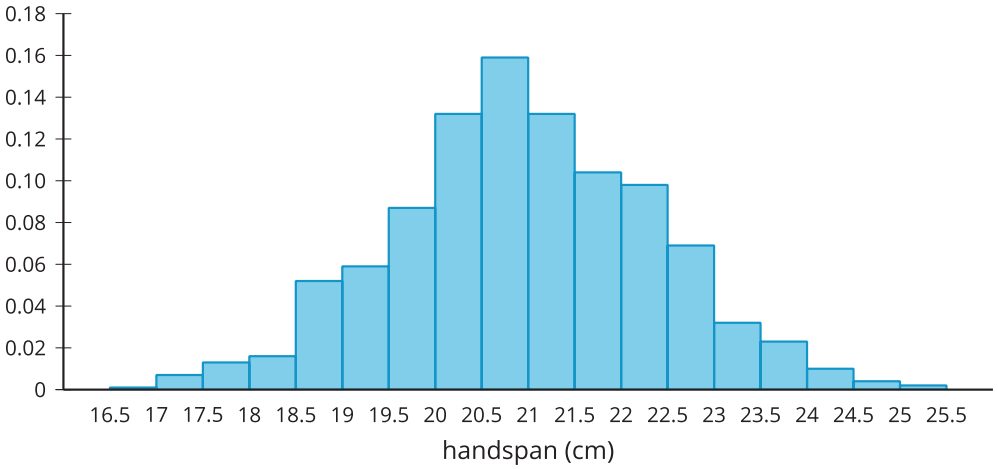
For example, a study measured the handspan, in centimeters, of 1,000 adults. A handspan is the distance from the thumb to the smallest finger when the fingers are stretched out as much as possible. A histogram shows the number of people whose handspans are in certain intervals. The height of each bar in a histogram represents the frequency for the corresponding interval.

In this example, there are 132 adults in the study whose handspans are at least 20 centimeters but less than 20.5 centimeters.




A histogram that uses the relative frequencies shows a distribution with the same shape, but the heights of the bars represent the relative frequency for the corresponding intervals.

For example, out of all the adults in the study, 13.2% have handspans that are at least 20 centimeters but less than 20.5 centimeters (since $\frac{132}{1,000} = 0.132$).




Similarly to how we can model data in a scatter plot with a line or other curve so that additional information can be estimated or predicted, it can be useful to model an approximately symmetric and bell-shaped distribution with a particular distribution called the *normal distribution*. A **normal distribution** is symmetric and bell-shaped, has an area of 1 between the $x$-axis and the curve, and has the $x$-axis as a horizontal asymptote. A normal distribution is determined entirely by the mean and standard deviation.

For the handspan data, the mean is 20.9 cm and the standard deviation is 1.41. The curve that represents the normal distribution with this same mean and standard deviation is shown on top of the relative frequency histogram for the actual data.

Notice that the curve does a fairly good job of modeling the actual data in this situation, although it is not perfect.

