Distribution Shapes

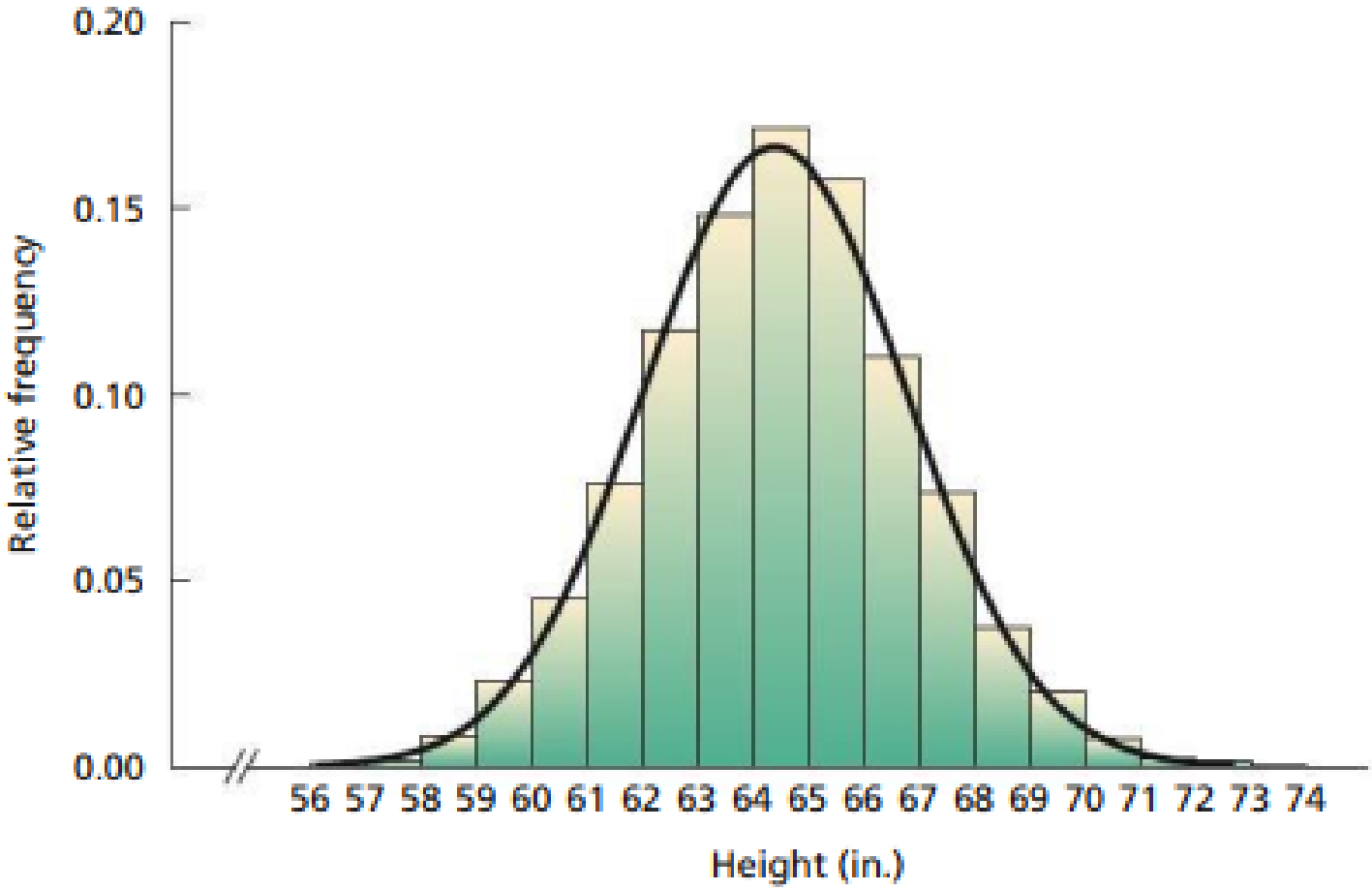
Def: The distribution of a data set is a table, graph, or formula that provides the values of the observations and how often they occur.

We've already shown distributions in the form of

- frequency distributions.
- frequency histograms.
- relative-frequency histograms.
- dot plots.
- stem-and-leaf diagrams.
- pie charts.
- bar charts.

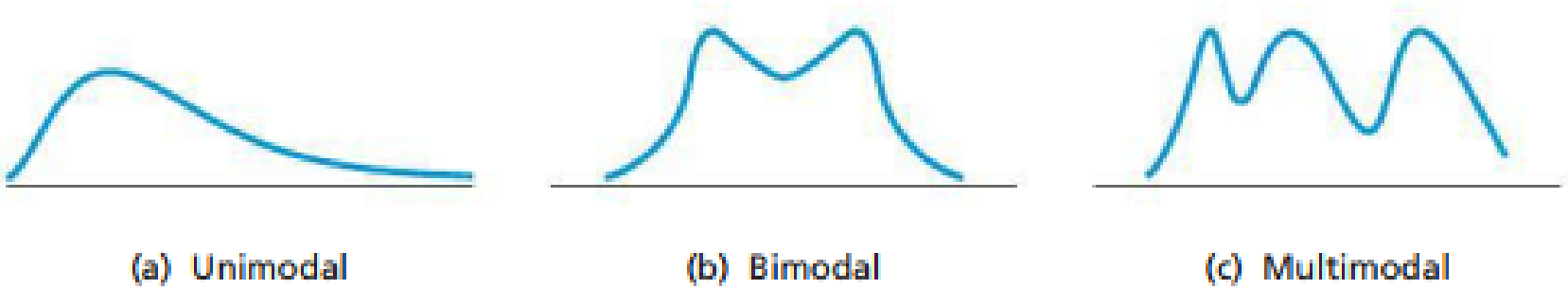
With quantitative data, we are particularly interested in the **shape** of a distribution.

As we go, this will help us choose methods for statistical analysis.



Relative frequency histogram of 3294 women at a midwestern college.

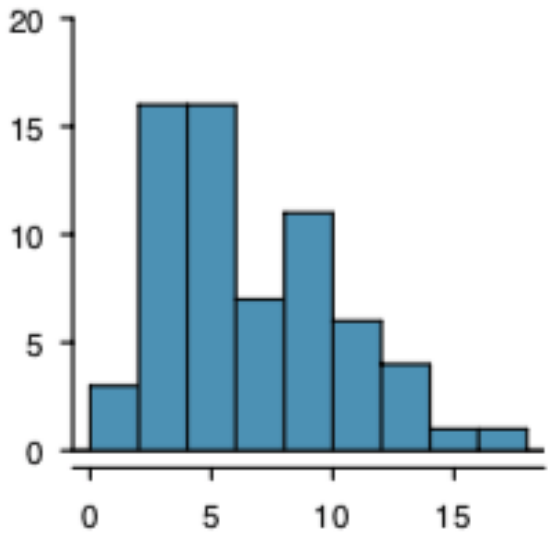
Modes

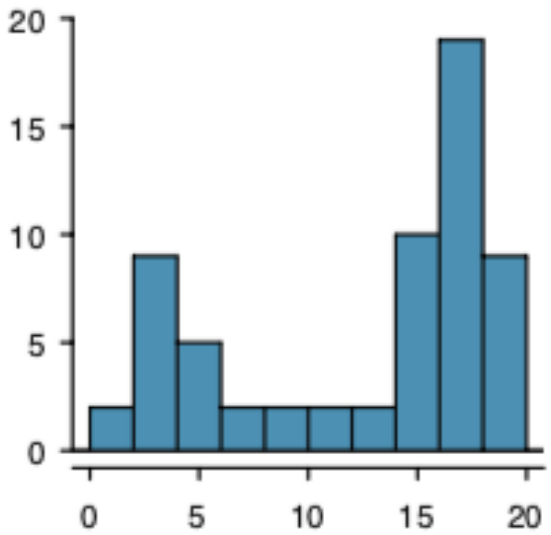


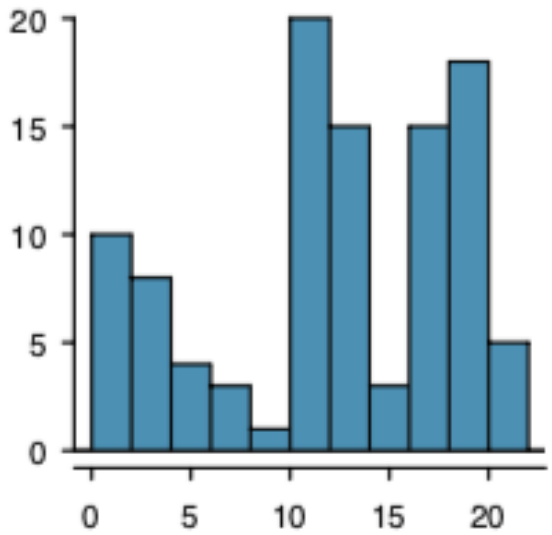
A **mode** is any prominent peak in the distribution.

- A distribution with one prominent peak is called **unimodal**.
- Distributions with two prominent peaks are **bimodal**.
- Distributions with three or more promiment peaks are **multimodal**.

Modes







How many modes are there in each distribution?

Modes

Bin widths, our particular sample, and differing opinions can all impact where we see a "prominent" mode.

...but that is okay! The goal of examining the shape of our data is simply to better understand the nature of our data. This allows us to make more informed technical decisions down the line.

Symmetry

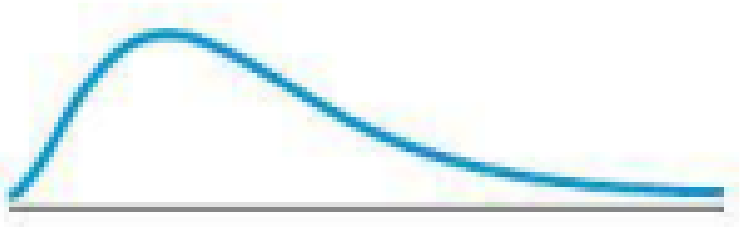
(a) Bell shaped

(b) Triangular

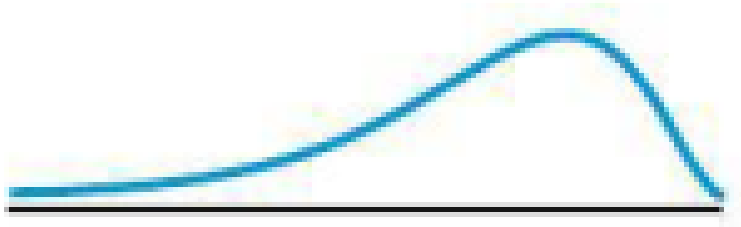
(c) Uniform (or rectangular)

- If the data have roughly equal tails, we say the distribution is **symmetric**.

Skewness



(a) Right skewed



(b) Left skewed

- If the data have a long, thin right tail, so we say that the shape is **right skewed**.
- If the data have a long, thin tail on the left, we say that the shape is **left skewed**.

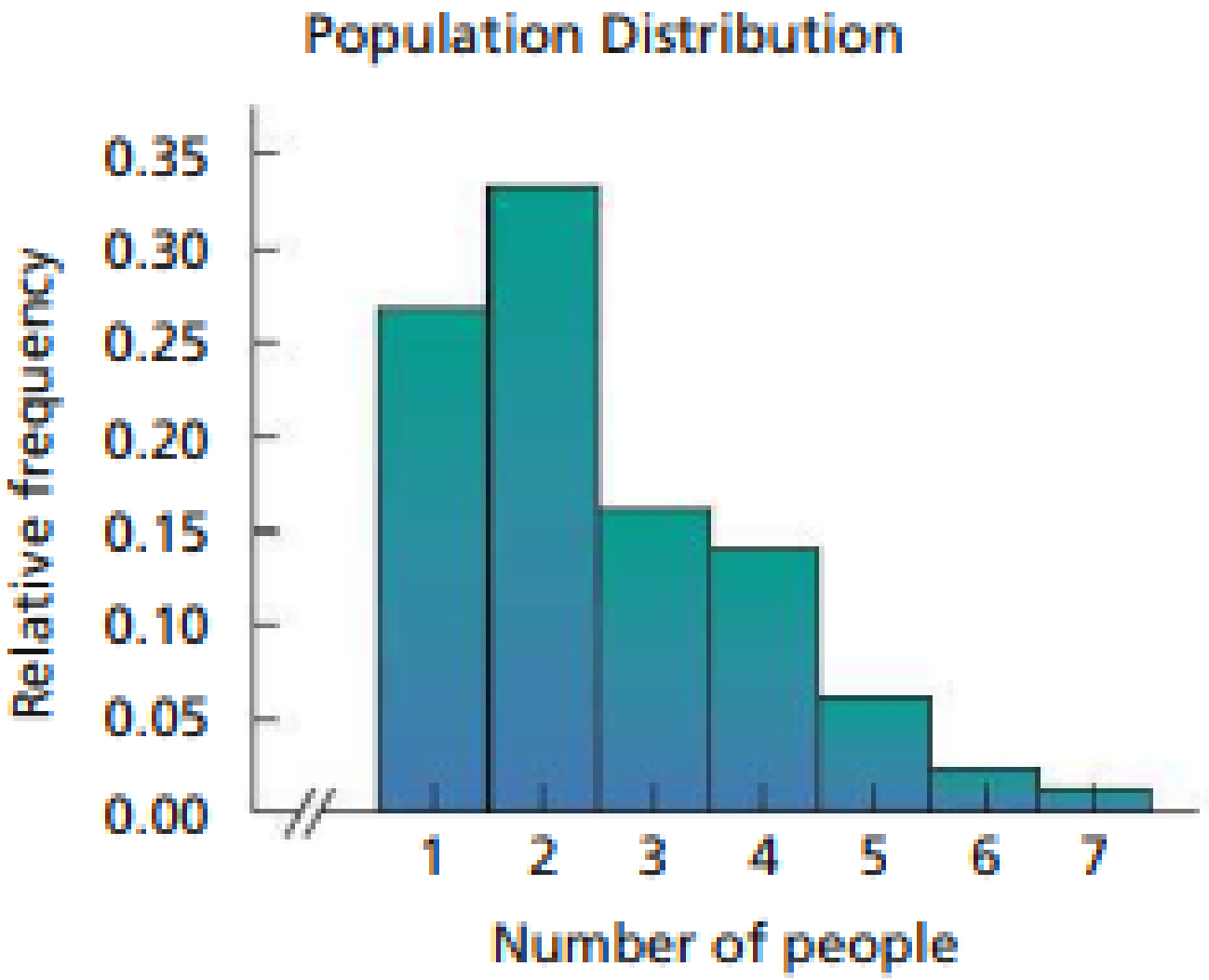
Def: Population data are the values of a variable for the entire population (also called census data).

Def: Sample data are the values of a variable for a sample of the population.

Def: The distribution of population data is called the population distribution or the distribution of the variable.

Def: The distribution of the sample data is called the sample distribution.

Population Distribution, Household Size



Six Sample Distributions (100 Households Per Sample)