Example 5-1: Candy Colors
a) Take a random sample of 25 candies and record the number and proportion of each color:

<table>
<thead>
<tr>
<th>Color</th>
<th>Number</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>orange</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yellow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>brown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b) Is the candy’s color a quantitative or a categorical variable?

c) Is the proportion of orange candies among the 25 that you selected a parameter or a statistic? What symbol should we use for it?

d) Is the proportion of orange candies manufactured by Hershey a parameter or a statistic? What symbol should we use for it?

e) Do you know the value of the proportion of orange candies manufactured by Hershey?

f) Do you know the value of the proportion of orange candies among the 25 that you selected?

g) Did every student obtain the same proportion of orange candies in his/her sample?

h) If every student was to estimate the population proportion of orange candies by the proportion of orange candies in his/her sample, would everyone arrive at the same estimate?

- The values of statistics vary from sample to sample. This phenomenon is called sampling variability. Fortunately, if we look at the results of many samples, there is a predictable pattern to this variability.

i) Add your sample proportion of orange candies to the graph on the board. Around what value (roughly) are the sample proportions centered?

- Since random sampling is unbiased, the actual value of the population proportion should be close to the center of these sample proportions.
j) If every student was to estimate the population proportion of orange candies by the proportion of orange candies in his/her sample, would most estimates be reasonably close to the true parameter value? Would some estimates be way off? Explain.

We need to take more samples to see the pattern of how sample statistics vary more clearly. For this we can turn to an applet called “Reese’s Pieces.” For now we will suppose that 45% of the population is orange.

k) Use the “Reese’s Pieces” applet to draw 500 samples of 25 candies each, assuming that the population proportion of orange is .45. (Pretend that this is really 500 students, each taking 25 candies and counting the number of orange ones.) Sketch and describe a graph of the sample proportions of orange obtained.

l) Is there an obvious pattern to the distribution of the sample proportions of orange candies? Is it approximately normal?

- Even though the sample proportion of orange candies varies from sample to sample, there is a recognizable long-term pattern to that variation. This pattern is called the sampling distribution of the statistic.

m) What are the mean and standard deviation of the sample proportions of orange candies?

n) Now assume that the population proportion of orange candies is .55. Again use the applet to draw 500 samples of 25 candies each. How has the distribution changed?

- shape:
- center:
- variability:
o) Now use the applet to draw 500 samples of 100 candies each (so these samples are four times larger than the ones you gathered in class). How has the distribution of sample proportions changed (or not changed) from when the sample size was only 25 candies?

- **shape:**
- **center:**
- **variability:**

- A larger sample size produces less variability in sample statistics.

**Key result:** Suppose that the proportion of a population having some characteristic is denoted by \( \pi \), and suppose that a random sample of size \( n \) is taken from the population. Then the sampling distribution of the sample proportion \( \hat{p} \) is approximately normal with mean \( \pi \) and standard deviation \( \sqrt{\frac{\pi(1-\pi)}{n}} \). This approximation is generally considered to be valid as long as \( n\pi \geq 10 \) and \( n(1-\pi) \geq 10 \).

**Example 5-2: Revisiting Previous Examples**

Reconsider again the simulation analyses that we have performed:

Infants’ toys \( (n = 16, \pi = .5) \)  \hspace{1cm} RPS \( (n = 119, \pi = 1/3) \)  \hspace{1cm} Racquet spin \( (n = 100, \pi = .5) \)

See what happens if we convert these graphs from number of successes to proportion of successes:
a) How do the shapes of these graphs compare between counts and proportions?

b) Comment on the center for each distribution of sample proportions. Explain why these make sense.

c) Comment on how the variability compares between the first and third graphs of sample proportions. Explain why this makes sense.

d) For the rock/paper/scissors study with 119 subjects, use the key result to describe how the sample proportion who choose “scissors” would vary from sample to sample if the null hypothesis (that people choose “scissors” with probability 1/3) were true. Also draw a sketch to represent this sampling distribution.

e) Recall that in the actual study, 14 of the 119 subjects chose “scissors.” Calculate the z-score for this result. Interpret what this value reveals.

f) What does this z-score reveal about the p-value? Explain.