Recall the important terms **population, sample, parameter, statistic**.
- A number describing a population is a **parameter**.
  - A parameter’s value is fixed but typically not known.
  - Symbols: \( \mu \) for popn mean, \( \sigma \) for popn std dev, \( \pi \) for popn proportion
- A number describing a sample is a **statistic**.
  - A statistic’s value can vary from sample to sample.
  - A statistic is often used to estimate a population parameter.
  - Symbols: \( \bar{x} \) for sample mean, \( s \) for sample std dev, \( \hat{p} \) for sample proportion

**Example 11-1: Candy Colors**
a) Take a random sample of 25 candies and record the number and proportion of each color:

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

b) What are the cases (observational units) here?

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

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

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

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

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

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

i) 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.
j) 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.

k) 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.

l) 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.

m) 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.

n) What are the mean and standard deviation of the 500 simulated sample proportions of orange candies?
o) Now assume that the population proportion of orange candies is .55. What do you think will change in the sampling distribution? Again use the applet to draw 500 samples of 25 candies each. How has the distribution changed?

- shape:

- center:

- variability:

p) 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), after predicting how you think the distribution will change. 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.

<table>
<thead>
<tr>
<th>Central Limit Theorem (CLT) for Sample Proportion:</th>
</tr>
</thead>
</table>
| 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 \( n(1-\pi) \geq 10 \).

q) Draw a sketch to represent this result, first in general and then applied to the candy example with \( n = 25 \) and assuming that the population proportion of orange is .45.
Example 11-2: Presidential Election
Among California voters in the 2008 election, 61% voted for President Obama.

a) Is this number a parameter or a statistic? Also identify it with the appropriate symbol.

Now suppose that you take a simple random sample of 200 California voters from 2008, and you determine the proportion of them who voted for Obama.

b) Is this number a parameter or a statistic? Also identify it with the appropriate symbol.

c) How would the sample proportion of Obama voters vary from sample to sample? Describe its shape, center, and spread. Also draw a well-labeled sketch to illustrate how this sample proportion would vary. Also check the conditions for whether the approximation is reasonable.

d) Between what two values would 95% of the sample proportions fall?

e) Determine the (approximate) probability that less than half of your simple random sample would have voted for Obama. (First shade the appropriate region in your sketch, then calculate the relevant z-score, and then use the normal probability table.) Also express this probability using symbols and in words.
f) Determine the (approximate) probability that more than two-thirds of your simple random sample would have voted for Obama.


g) Determine the (approximate) probability that the proportion of Obama voters in your simple random sample would be between .57 and .65.


h) Now suppose that you decide instead to take a simple random sample of 500 California voters. Without doing any probability calculations, describe how your answers to e), f), and g) will change. Explain your reasoning.


i) Perform the probability calculations for e), f), and g) with a sample size of 500. Comment on how these probabilities have changed with the larger sample size.