Now that you know the CLT, we can make a cool connection with something you learned in your prerequisite course. Let me first introduce the Bernoulli distribution.....

**Bernoulli distribution**

We say that the random variable \( Y \) has the Bernoulli distribution—that is, \( Y \sim Bernoulli(\pi) \)—if \( Y \) takes only values 0 and 1, and the probability that \( Y \) takes value 1 is \( \pi \). In practical applications, the value 1 corresponds to a “success,” and 0 to a “failure,” so that the parameter \( \pi \) is interpreted as the probability of success for what we call a Bernoulli trial. **Note: we are talking about one trial, not multiple trials, and the result is coded 0 or 1.** The PMF, mean, and variance of a Bernoulli random variable are given below, but note that once I give you the PMF, you should be able to derive the mean and variance using the definitions of these quantities for discrete random variables (though it isn’t easy—you sometimes have to use some nifty mathematical tricks).

\[
P(Y = y) = \pi^y (1 - \pi)^{1-y} = \begin{cases} \pi, & y = 1 \\ 1 - \pi, & y = 0 \end{cases}
\]

\[
EY = \pi
\]

\[
V(Y) = \pi(1 - \pi)
\]

You are probably already realizing there is a relationship between the Bernoulli distribution and the binomial distribution.

**The binomial distribution:**

Suppose we have \( n \) independent Bernoulli trials, each with the same probability of success \( \pi \). In other words, suppose we have a random sample from the Bernoulli distribution: \( Y_1, \ldots, Y_n \overset{iid}{\sim} Bernoulli(\pi) \). Let \( X = \sum_{i=1}^{n} Y_i \) be the sample total, which can be interpreted as the number of successes: \( X = \sum_{i=1}^{n} Y_i \). You probably recognize this as a binomial random variable: \( X \sim bin(n, \pi) \).

The PMF is given by:

\[
P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} = \frac{n!}{x!(n-x)!}
\]
Although you could derive the expected value and variance using the definitions of expected value and variance, or you could look them up in a book or online, down below I want you to use the theorems I’ve taught you to derive the expected value and variance of the binomial distribution from those of the Bernoulli, by recognizing that $X$ is the total of a random sample of $Y$‘s. (This is much quicker than using the definition of expected value and variance, trust me! You have to have some cool summation tricks up your sleeve to do it the other way.)

1. Derivation of $EX$ from recognizing that $X$ is the total from a random sample from the Bernoulli distribution:

$$EX = E\left(\sum_{i=1}^{n} Y_i\right) = n\mu_Y = n\pi$$

2. Derivation of $V(X)$:

$$VarX = Var\left(\sum_{i=1}^{n} Y_i\right) = n\sigma_Y^2 = n\pi(1-\pi)$$

The normal approximation to the binomial distribution—justified by the CLT (an important theorem):

It is very difficult for even your computer or calculator to compute factorials of large integers. That means that if $n$ is large, calculating binomial probabilities is difficult. But since the binomial random variable is really the sample total from a random sample of Bernoulli random variables, what does the Central Limit Theorem—did I mention how important this theorem is?—tell us about an approximation we can use to calculate binomial probabilities? You may have been taught this in your prerequisite course.

By the CLT, a linear combination of iid RV’s is normally distributed, so long as no one of the $X_i$ has a disproportionally large weight. Thus, for large $n$

$$X \xrightarrow{\text{approx}} N\{n\pi, n\pi(1-\pi)\}.$$

How large must $n$ be in this particular application of the CLT?

As always, that depends on the shape of the original distribution, in this case, the distribution of the $Y_i$.

For the Bernoulli distribution, when $\pi = .5$, and thus $1-\pi = .5$ the PMF is perfectly symmetric—in fact, it is a uniform distribution. When $\pi$ gets further away from .5, the PMF becomes more skew. See the graphs below for a.) $\pi = .4$ and b.) $\pi = .1$. 
Rule of thumb for how large \( n \) should be for the normal approximation to the binomial distribution:

If both of the following are true, then the sample size is large enough to overcome any skewness that might exist in the original Bernoulli PMF:

1. \( n\pi \geq 10 \), and
2. \( n(1 - \pi) \geq 10 \)

Bead box activity:

Estimating population probability of success, \( \pi \), also called population proportion of “successful” outcomes.

We will use the sample proportion of successes, \( \bar{X} / n \), to estimate the probability of success \( \pi \).

Theoretical result:

You can use either the CLT or the theorems about linear combinations of RV’s to derive the sampling distribution of \( \bar{X} / n \) from the information above. Write that below:

\[
\text{If } X \overset{\text{approx}}{\sim} N\left(n\pi, n\pi(1-\pi)\right), \text{ then } \frac{X}{n} \overset{\text{approx}}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)
\]

Seeing the theoretical result in action:

Three bead boxes:

1. **John Deere**: yellow and green beads—**green** is success
2. **Christmas**: green and red beads—**green** is success
3. **NC State**: red and white beads—**white** is success.
Three sample sizes (three paddle sizes):
1. \( n = 30 \)
2. \( n = 50 \)
3. \( n = 100 \)

What to do:
1. Each person in your group should “sample” from the box **with each paddle**.
   1.1. Make sure to “stir” gently between samples.
   1.2. Make sure all beads are removed from a paddle each time (you may need to tap it **gently** on the side of the box).
2. Put a dot above the sample proportion on the chart paper for the appropriate sample size.
3. Record the sample proportion in the table, and/or put a hash mark noting that a value was repeated. This will allow us to store the data for later use.
4. Keep sampling until I tell you to switch stations. Let everybody have a turn. Fun, fun, fun!
5. Don’t spill the beads and mess up the population!

Guess the proportions:

<table>
<thead>
<tr>
<th>Population</th>
<th>Your estimate of ( \pi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Deere</td>
<td>.2</td>
</tr>
<tr>
<td>Christmas</td>
<td>.5</td>
</tr>
<tr>
<td>NCSU</td>
<td>.1</td>
</tr>
</tbody>
</table>