Math 247: The Central Limit Theorem for Sample Proportions (Section 7.3)
and Introduction to Hypothesis Testing (Section 8.1)

The goal here is to finalize our observations about the Sampling Distribution for $\hat{p}$, and turn the Sampling Distribution into a tool for making inferences about a population, using sample data.

Remember, since we rarely have census data, we have to INFER what's true about a population by using sample data!

The Central Limit Theorem: The Sampling Distribution of $\hat{p}$ **, will be approximately a Normal Distribution** as long as the following conditions are met:

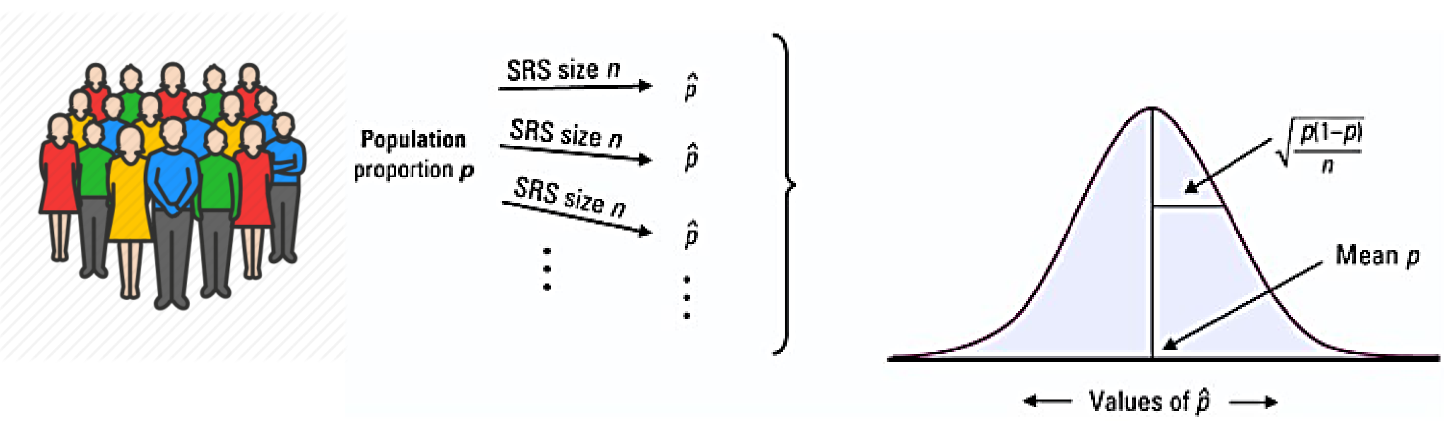
Conditions:

1. **Random and Independent:** We must have a random sample from a population and the observations must be independent from one another.

2. **Large Sample:** The sample size is large enough to make the distribution approximately normal.

   How to check this: **Expected Count for each group must be at least 10**:
   Success group: $E = np$ (if p is unknown, use $E = n\hat{p}$)
   Failure group: $E = n(1-p)$ (if p is unknown, use $E = n(1-\hat{p})$)

3. **Large Population:** The population is at least 10 times greater than the sample. This is because we're *sampling without replacement* and the large sample from a small population would change the proportions!



The center of the distribution is $\mu_{\hat{p}} = p$ the *population proportion*	The spread of the distribution, as always, is given by the standard deviation $\sigma_{\hat{p}} = SE = \sqrt{\frac{p(1-p)}{n}}$ which is this context is called the "standard error" (SE)
The z-score, as always, measures how many standard deviations an observed value is from the mean.	$z = \frac{\text{observed} - \text{mean}}{\text{SD}} = \frac{\hat{p} - p}{SE}$

Example: **Is poverty increasing in SLO?** The U.S. Constitution requires the government to conduct a census every 10 years. (This cost about $15 billion in 2010!) In the last census (year 2010), the data showed 12.8% of the residents in SLO county were living below the poverty line. In 2016, the American Community Survey (also conducted by the U.S. Census Bureau) found that in a sample of 3000 residents in SLO county, 426 (14.2%) were below the poverty line. *(source: http://www.slohealthcounts.org/indicators/index/view?indicatorId=347&localeId=277)*

Since the proportion (14.2%) didn't come from census data, could this apparent increase just be the result of sampling variability, i.e., just due to chance? Or, is this a *statistically significant* increase in the proportion of people living in poverty? How can we tell?

To answer this, we can find the probability that 14.2% or more of the people in the sample would be living in poverty **if** the true population proportion was actually still 12.8%.

We can organize this analysis into a Hypothesis Test

Step 1: Hypothesize

Assume, for argument's sake, that nothing has changed in SLO; i.e., that poverty levels are the same as before.

This assumption (the "**IF**" above) is the "Null Hypothesis" and is denoted H_O.
Write the Null hypothesis in words and in symbols.

H_O

What we're wondering, though, is whether poverty has increased. We form the "Alternative Hypothesis" based on whether we suspect an increase (or in other problems, a decrease, or just a change.)

Write the Alternative Hypothesis in words and symbols.

H_a

So, here's the question: Does the sample data provide evidence that poverty levels have increased?

Step 2: Prepare (Choose test, set alpha, check conditions)

The test we're going to use here is called the "One Proportion z-Test".

This is where the **Central Limit Theorem** comes in. We have to check that the conditions to use the Central Limit Theorem are satisfied by our sample.

Check that the Conditions are satisfied

a) **Random sample and Independent observations?**

b) **Large Sample?**
Note: We have defined "success" and "failure" by how we've written our hypotheses.

Success group:

Check: $E = np \geq 10$?

Failure group:

Check: $E = n(1 - p) \geq 10$?

c) **Large Population?** $Pop \geq 10 \cdot n$

Step 3: Compute

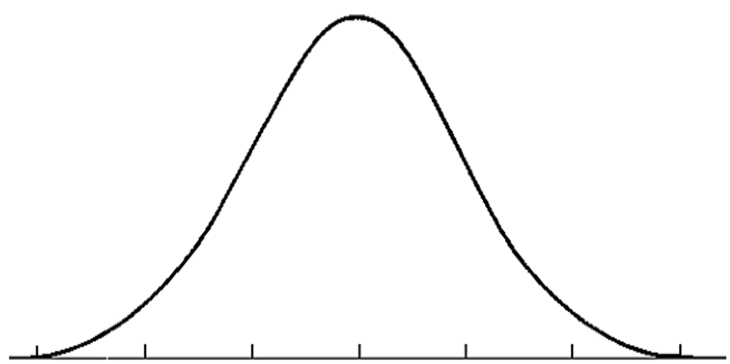
Now we're going to find the probability that we would get an observation as or more extreme as we did IF the null hypothesis was actually true.

In this problem, we could state this as "If poverty levels now were still at the same level as in 2010 (12.8%), what are the chances we would get a sample of 3000 people that had 426 people in poverty; i.e., showed a poverty level of 14.2% or more?

Parking Lot:

- Draw a normal curve that represents the Sampling Distribution of $\hat{p}$. The null is the center!
- Use the Standard Error formula to find the standard deviation of the sampling distribution.

$$SE = \sqrt{\frac{p(1-p)}{n}}$$



- Plot the Sample Proportion
- Write a probability problem in the proper notation and shade in the area that represents the probability

- Find the z-score

Test statistic: $z = \frac{\text{observed} - \text{center}}{\text{SD}} = \frac{\hat{p} - p}{SE}$

- Use the Normal Distribution Calculator to find the probability. Use the proper notation to express the probability.

Step 4: Interpret