Once you have a sample, what can you say about the *population* proportion? How confident are you of the result?

General Education Statistics
Class Notes
Confidence Intervals for Population Proportions (Section 9.1)

Remember the point of taking a sample is to reflect back on the population. Say that we find 22.4% of a sample has tried and like fried Twinkies. But what does that say about the larger population?

We will round proportions to three decimal places (or the nearest tenth of a percent).

Definition: A **point estimate** is the value of a (sample) statistic that estimates the value of a (population) parameter. For instance, the sample proportion we have been working with, $\hat{p} = x/n$ where x is the number of individuals in the sample that share some characteristic and n is the sample size, is a **point estimate**.

We could simply transfer this point estimate to the population. For instance, we could say that 22.4% of Americans have tried and like fried Twinkies. However, since we know this sample is just one of many possible samples and they vary because of their random nature, we now look at a method that gives us more accuracy, confidence intervals.

Definition: A **confidence interval** for an unknown parameter consists of an interval of numbers based on a point estimate.

The **level of confidence** represents the expected proportion of intervals that will contain the parameter if a large number of different samples are obtained. The level of confidence is denoted $(1 - \alpha) \cdot 100\%$.

Confidence level	**Value of α**
90%	0.10
95%	0.05
99%	0.01
etc.	

Confidence interval estimates for the population proportion are of the form

point estimate ± margin of error.

The **margin of error** is a measure of how accurate the point estimate is.

The foundation of what we do here is based on the fact that the sampling distribution of $\hat{p}$ is normal. Without getting into too many details, let's get started.

expl 1: In a random sample of 125 Americans, 28 of them have tried and like fried Twinkies. We have previously found the sample proportion to be $\hat{p} = .224$. The margin of error for a 95% confidence interval is .073. Form the confidence interval using the formula
point estimate ± margin of error.

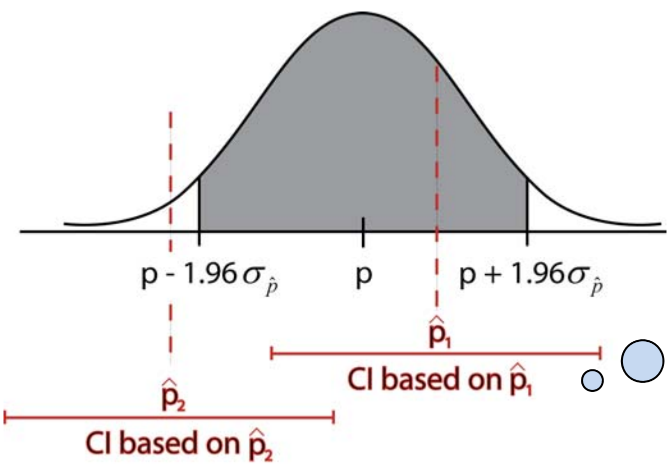
We will learn how to find the margin of error later.

Complete the sentence:

We are 95% confident that the true proportion of Americans who have tried and like fried Twinkies is between _________ and _________.

What does this confidence mean?

These confidence intervals are based on the fact that the sampling distribution of $\hat{p}$ is normal. If we picture a normal curve with the population parameter in the middle and plus or minus 1.96 standard deviations marked, we see where 95% of all possible sample proportions should lie (shaded region below). [The Empirical Rule, which tells us that 95% of all observations lie within two standard deviations, has rounded 1.96 to 2.]



Imagine two different samples, each yielding its own $\hat{p}$. Form confidence intervals (abbreviated CI) for each $\hat{p}$.

Most confidence intervals (in fact, 95% of them) will contain the population proportion. However, a few samples (only 5%) will yield a sample proportion so far from the true mean that the confidence interval will *not* contain the true population proportion.

The confidence level indicates that *the method we are using* will yield a confidence interval that indeed does contain the true population proportion 95% of the time.

In practice, only one sample is taken and only one confidence interval is formed.

> Margin of error determines the length of the interval. This is *not* the same as the confidence level.

Margin of Error:

The margin of error depends on three factors.
They are level of confidence, sample size, and the standard deviation of the population. Let's investigate this a bit.

expl 2a: From example 1, we have the 95% confidence interval for the percentage of Americans who have tried and like fried Twinkies (between 15.1% and 29.7%). If we want to be 99% confident that the true percentage lies in our interval, would we need a shorter or longer interval?

expl 2b: The margin of error measures the variability of the sample. If we sample more people, the Law of Large Numbers says we will *decrease* the variability. Would this result in a shorter or longer interval?

> We can also see this in the formula below.

Constructing a $(1-\alpha)\cdot 100\%$ Confidence Interval for a Population Proportion:

Suppose that a simple random sample of size n is taken from a population. A $(1-\alpha)\cdot 100\%$ confidence interval for p is given by the following quantities

Lower bound: $\hat{p} - z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Upper bound: $\hat{p} + z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

> This craziness we add or subtract is the **margin of error**. It is sometimes denoted as E.

The **critical value**, $z_{\alpha/2}$, is based on the confidence level and can be found in the table below.

> These are common but any level of confidence is possible.

Critical Values for Confidence Intervals	
Level of Confidence $(1-\alpha)\cdot 100\%$	**Critical Value**
90%	1.645
95%	1.96
99%	2.575

The **critical value** is how many standard deviations from the population mean the sample statistic can be and still result in a confidence interval that includes the parameter with the desired level of confidence.

expl 3: A survey of 2306 adult Americans aged 18 and older conducted by Harris Interactive found that 417 have donated blood in the past two years. Answer the following questions.

a.) Calculate the point estimate for the proportion of the population who have donated blood in the past two years. Round to three decimal places.

b.) Verify that the requirements for constructing a confidence interval about p are satisfied.

Check that $n \le .05N$ and $n\hat{p}(1-\hat{p}) \ge 10$.

We use $\hat{p}$ here instead of p because we do *not* know p.

c.) Find $z_{\alpha/2}$ in the table on the previous page for a 90% confidence level. Calculate the margin of error using the formula $z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

d.) Construct a 90% confidence interval for the population proportion of adult Americans who have donated blood in the past two years. Complete the sentence below.

We are _____ confident that the true proportion of adult Americans who have donated blood in the past two years is between ________ and ________.

This does *not* mean that the *probability* the true percentage is in this interval is 90%.

Using Technology:

We will redo example 3 using StatCrunch and the TI calculators.

expl 4: A survey of 2306 adult Americans aged 18 and older conducted by Harris Interactive found that 417 have donated blood in the past two years. Construct a 90% confidence interval for the population proportion of adult Americans who have donated blood in the past two years.

Steps for TI Calculator:

> Technology does *not* check whether or not $n \leq .05N$ and $n\hat{p}(1-\hat{p}) \geq 10$.

Press **STAT** and arrow over to **TESTS**. Arrow down the list to find **A:1-PropZInt…**

You will enter the value for x, n, and the confidence level.

Select **Calculate** and press **ENTER**. It will calculate the interval, $\hat{p}$, and verify the sample size.

Steps for StatCrunch:

If you have raw data, enter it in the spreadsheet. If not, go on to the next step.

Then select **Stat** > **Proportion Stats** > **One Sample** > **With Summary** (or **With Data** if you entered data).

Enter the number of successes and observations. Under **Perform**, select **Confidence interval for *p***. Enter the desired **Level** of confidence and tell it you want the **Standard-Wald Method**. (If you entered raw data, select the column with your data.)

Click **Compute!** You can check your values for x and n. It tells you $\hat{p}$ and its standard deviation (**Std. Err.**). The interval is given as **L. Limit** and **U. Limit** on the far right of the table output.

Did the technology match what we did by hand?

Calculating the critical value $z_{\alpha/2}$ when it's not in the table

This may happen because of the somewhat random numbers given in MSL Homework. We understand $z_{\alpha/2}$ to be the z-score such that the area under the normal curve to the right of this value is $\alpha/2$. Recall also that α is the value that defines the confidence interval as $(1-\alpha)\cdot 100\%$. First, figure out what α is and then use the Normal table or technology (InvNorm anyone?) to find the desired z-score.

expl 5: What is the critical value for a 94% confidence interval? Follow the steps.
a.) So, we know $1-\alpha = 0.94$. Solve for α.

b.) We need the critical value $z_{\alpha/2}$, the z-score such that the area under the normal curve to the right of this value is $\alpha/2$. What is the value of $\alpha/2$?

c.) Use technology or use the normal table to find the critical value.

> Your calculator may require you know the area under the curve to the *left* of this value. However, the critical value will always be positive.

Calculating Sample Size Needed for Confidence Intervals:

Before we sample, can we determine the sample size needed to then be able to form a 95% confidence interval with a given margin of error? You betcha!

The **sample size needed for a specified margin of error**, E, and level of confidence $(1-\alpha)$ is given by $n = \hat{p}\left(1-\hat{p}\right)\left(\frac{z_{\alpha/2}}{E}\right)^2$. This, you will notice, assumes you have a previous sample with a $\hat{p}$ value. That is *not* always the case.

Use E in decimal form.

If we have *no* previous value to use for $\hat{p}$, then we will use the simpler formula

$$n = 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2.$$

Always round the sample size value *up*.

expl 6: A television sports channel wants to estimate the proportion of Americans who watch professional football. What sample size should they use if they want to be within three percentage points with 95% confidence if …

a.) …they use a previous sample that showed 65% of Americans watch pro football?

Use E and $\hat{p}$ in decimal form.

b.) … they do *not* have a previous estimate for p?

Remember to round your final answers *up*.

Confidence Intervals and Overlap:
What does it mean when related confidence intervals overlap or do not? Can we glean information from it?

expl 7: A CBS/Times poll taken on October 28 – November 1, 2016 asked for whom 1,333 likely voters were going to vote. The results showed 47% for Hillary Clinton and 44% for Donald Trump. The margin of error for both point estimates was 2.7%. Answer the following questions. (source: https://elections.huffingtonpost.com/pollster/2016-general-election-trump-vs-clinton)

a.) Form a 95% confidence interval for the proportion of likely voters who will vote for Hillary Clinton.

b.) Form a 95% confidence interval for the proportion of likely voters who will vote for Donald Trump.

c.) Polls like this one were touted as showing Hillary Clinton in the lead. However, notice the two intervals overlap. What does this say about her "lead"?

Worksheet: Confidence Intervals and the sample size:
This worksheet shows how the sample size affects the interval length. It practices finding the margins of error and forming the intervals.

Optional Worksheet: Confidence level and interval length:
This worksheet will show how the confidence level (95% versus 99% versus 90%) affects the interval length. Again, it gives practice finding the margins of error and forming the intervals.