Week Seven: Confidence Interval (CI) about p

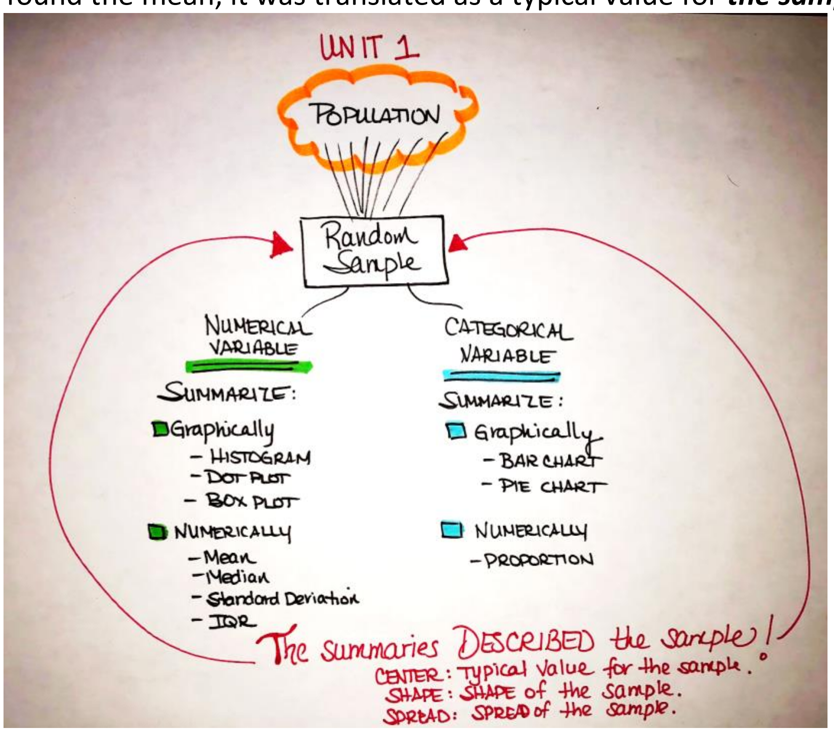
Week Seven Goals

- Introduction
- Sampling Distributions about the Population Proportion (p)
- Summary of STAT 145 So Far
- Confidence Interval about p, the population proportion
- Interpreting a Confidence Interval
- Meeting the Assumptions/Requirements for a CI about p
- Answering follow-up questions about the CI
- Building a Confidence Interval about p –WITH StatHelper
- Building a Confidence Interval about p –WITH MINITAB 19
- Working with an Existing CI
- Understanding How a CI changes
- Estimating Sample Size with Categorical Data

NOTE: If you choose to read the e-text as a resource for this section, you will notice that our e-text denotes the sample proportion with p' (as opposed to p-hat, $\hat{p}$).

Introduction

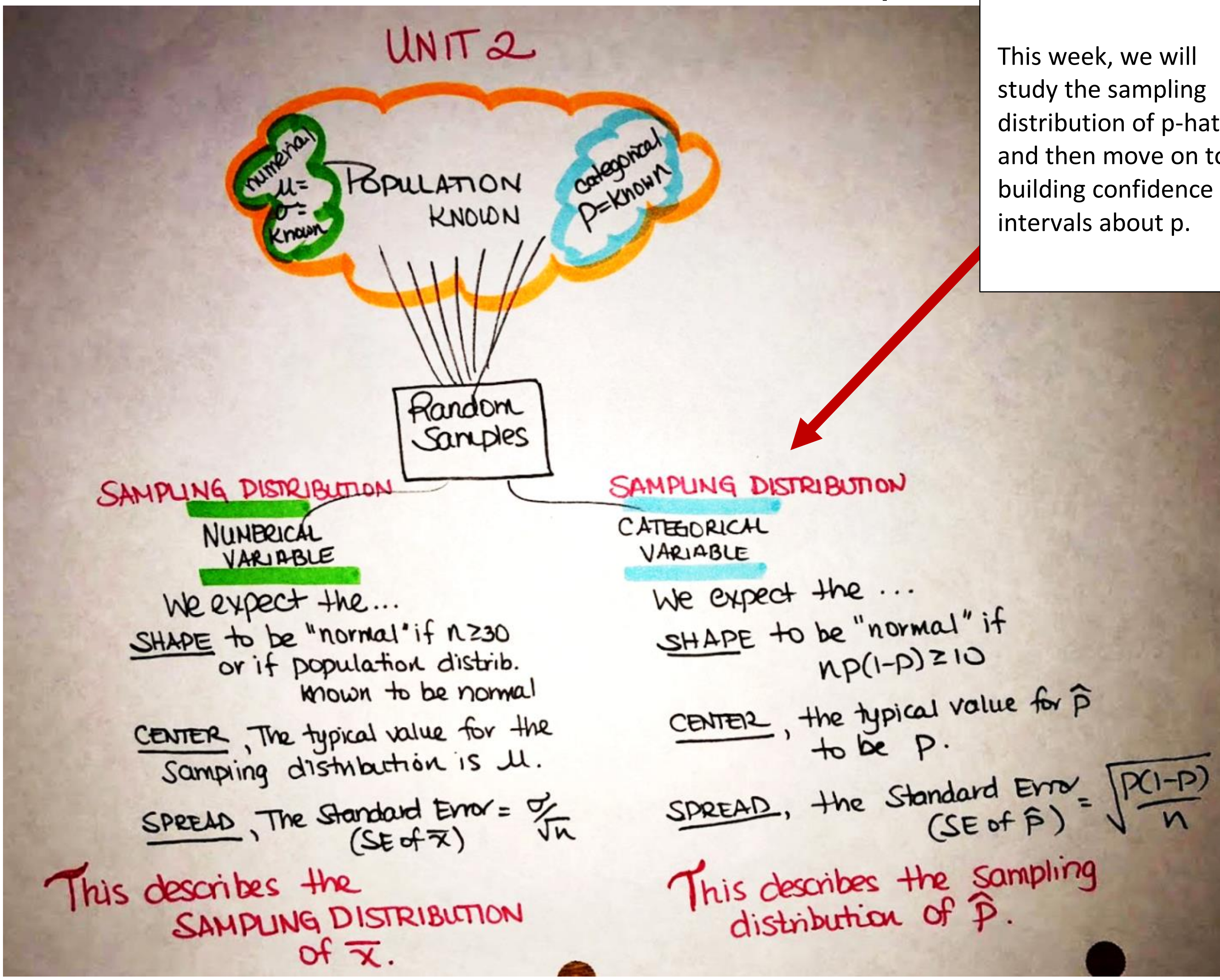
In Unit 1, we learned how to summarize data that were taken from a population. The summaries were ABOUT THE SAMPLE. When we found the mean, it was translated as a typical value for ***the sample.***



In week five, we learned how to describe and study a sampling distribution of $\bar{x}$ when we know information about the population. Knowing this information, we were able to answer probability questions.

Last week, in week 6, you learned how to build confidence intervals about the true mean. In both weeks, we were working with NUMERICAL data.

Now, let's work with CATEGORICAL data and describe sampling distributions of $\hat{p}$.



This week, we will study the sampling distribution of p-hat and then move on to building confidence intervals about p.

Sampling Distributions of $\hat{p}$

The distribution of the values of the sample proportions ($\hat{p}$, p-hat) in repeated samples (of the same size) is called the sampling distribution of $\hat{p}$.

To describe the sampling distribution of p-hat, we need to describe SHAPE, CENTER and SPREAD.

- **The SHAPE of the sampling distribution of p-hat can be assumed normal if n(p)(1-p) $\geq$ 10.**

Statisticians know that the sample size (n) needs to be large enough for the sampling distribution that has p at its center. To this end, different textbooks use different methods for determining what sample size is large enough. We will use the formula: **n(p)(1-p) $\geq$ 10**. If the product of n(p) (1 – p) is 10 or larger, then we can say that the sample size is large enough to assume a normal distribution. *NOTE* I am asking you to use THIS METHOD of checking whether the sample size is large enough when working with proportions; I am aware that our textbook does not use this formula.

- **The CENTER of the sampling distribution of p-hat is the population proportion (p), given in the problem.**

It is reasonable to expect all the sample proportions in repeated random samples to average out to the underlying population proportion, 0.6. In other words, the mean of the distribution of p-hat should be p.

- **The SPREAD of the sampling distribution of p-hat is $\sqrt{\frac{p(1-p)}{n}}$.**

Statisticians know that sample size plays a role in the spread of the distribution of sample proportion: there should be less spread for larger samples, more spread for smaller samples. In fact, the standard deviation of all sample proportions is directly related to the sample size, n as indicated by the formula

$$Standard\ Error = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

This website provides a nice explanation of shape, center and spread and was referenced.

Once we have established that the sampling distribution can be assumed normal (SHAPE) and you know the CENTER of the sampling distribution (it will be the population proportion) and SPREAD, we can work with the standard normal (Z) curve to answer probability questions.

Video Describing the Sampling Distribution for p-hat with Example from Prof. Coffey

Here is a video describing sampling distributions for p-hat: https://youtu.be/Nkuz-0oAUNg

Here is the same video with interpreting: https://youtu.be/5SYVoDpP2Dw

Example 1

Suppose the proportion of all college students who have used marijuana in the past 6 months is p = .40. For a random sample of size n = 200, representative of all college students on use of marijuana, describe the sampling distribution of $\hat{p}$ (p-hat) values.

A. Describe the SHAPE:

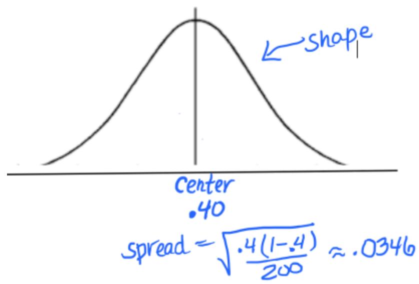
   Calculate (200)(.40)(1 - .40) = 48. Since this is greater than or equal to 10, we can assume a normal model.

B. Describe the CENTER:

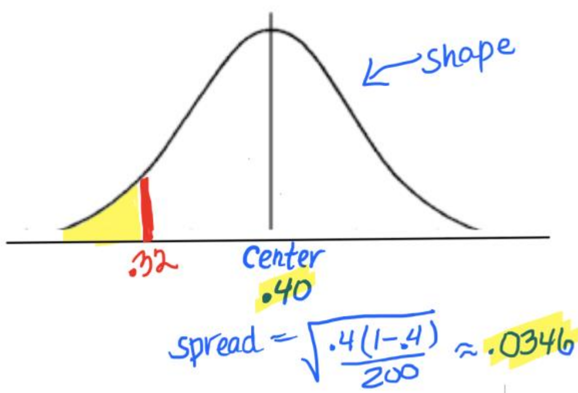
   The center of the sampling distribution (or typical value for p-hat) is .40, the population proportion.

C. Describe the SPREAD:

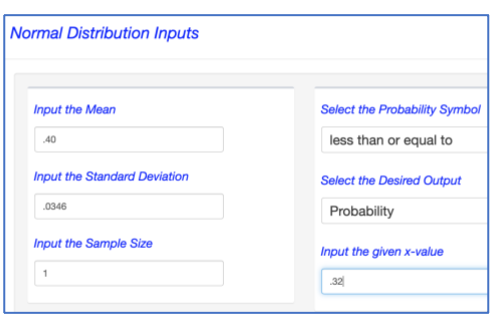
   Calculate the standard error for p-hat values by... $\sqrt{\frac{(p)(1-p)}{n}}$ $\sqrt{\frac{(.40)(1-.40)}{200}} = 0.0346$

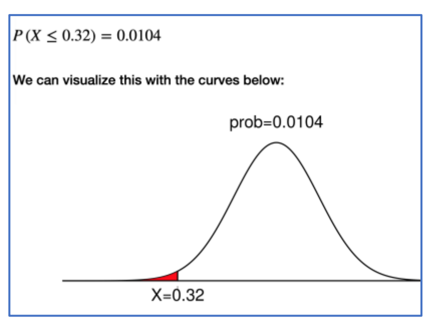


D. Now that we have established shape, center and spread, what is the probability that the sample proportion of students who have used marijuana in the past 6 months is less than .32? **ANSWER: 0.0104**



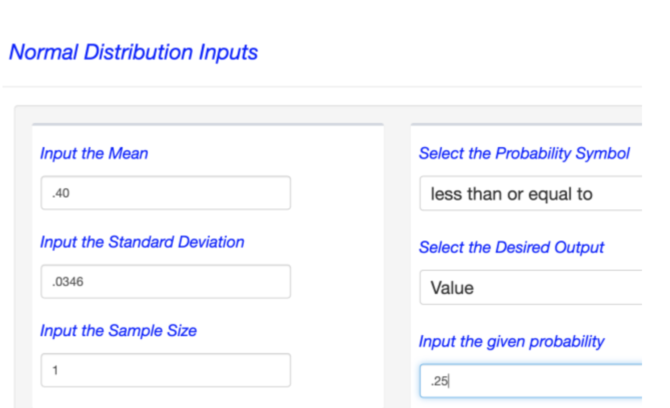
StatHelper: Choose Probability Distributions→Normal Distribution

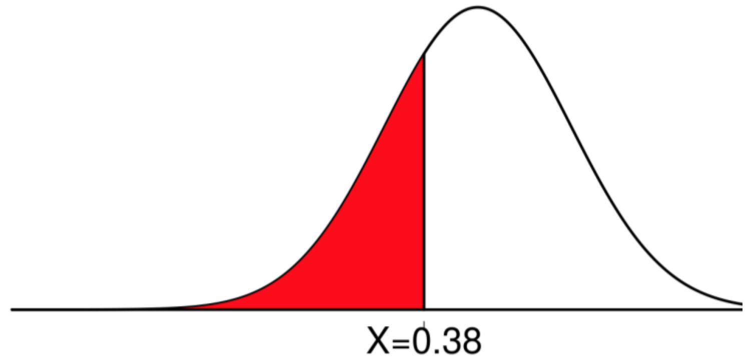




Minitab 19: Choose Calc→Probability Distributions→NORMAL; mean = .40, standard deviation = .0346

E. Now that we have established shape, center and spread, find the sample proportion of students who have used marijuana in the past 6 months that represents the 25th percentile? **ANSWER: .38**



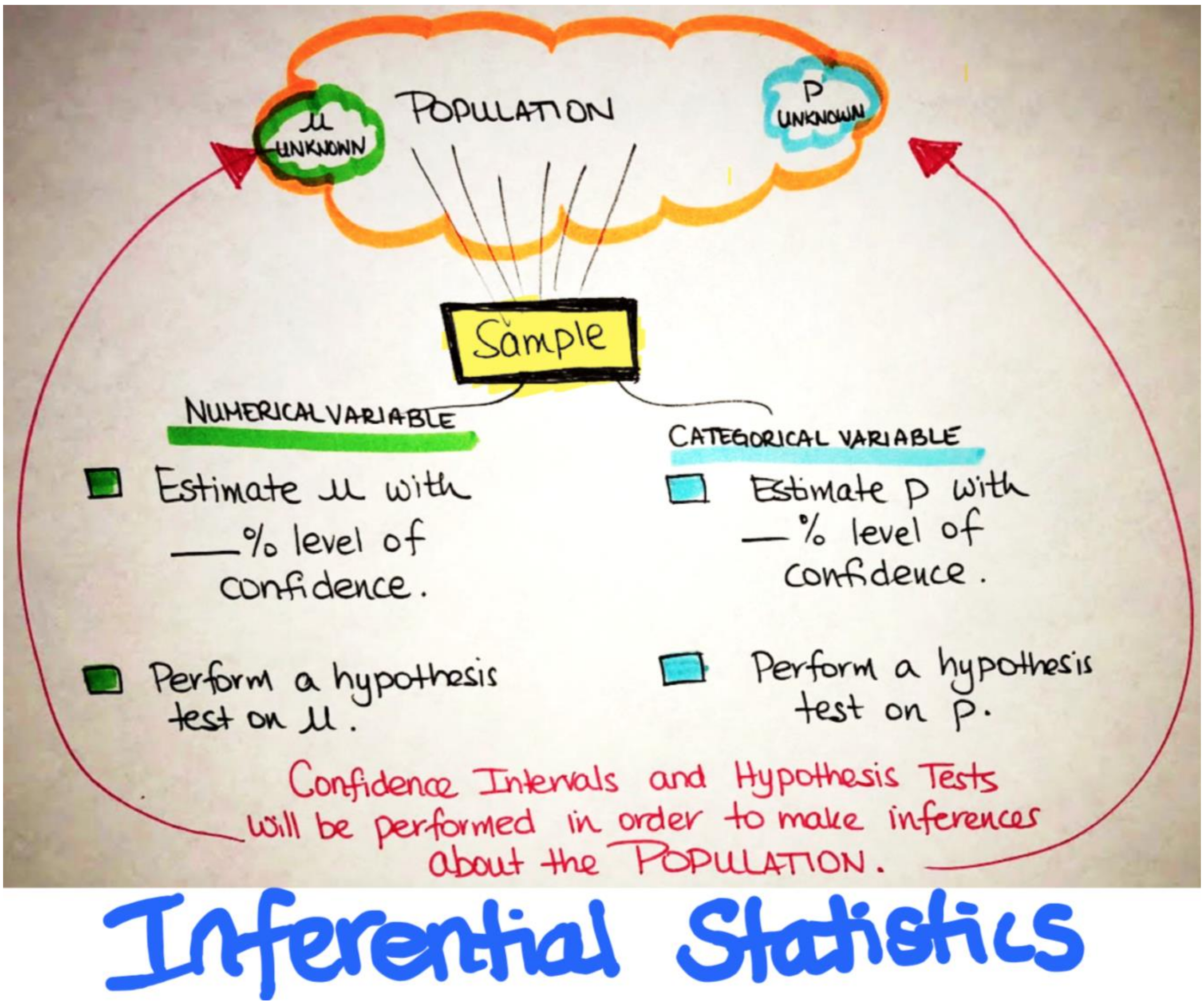


Summary of STAT 145 Class So Far

Video Summarizing this Statistics Course so Far from Prof. Coffey

Here is a video summarizing this course: https://youtu.be/ZULKVm40YNo

Here is the same video with interpreting: https://youtu.be/PbuJ-zqxCFk



NOTE: If you choose to read the e-text as a resource for this section, you will notice that our e-text denotes the sample proportion with p' (as opposed to p-hat, $\hat{p}$).

Confidence Interval (CI) about p

Now that we know how to describe the sampling distribution of $\hat{p}$ we will use this knowledge to begin making inferences about p, the population proportion.

Last week, when we built confidence intervals about mu, we were using the sample mean, $\bar{x}$, to make an inference about the population mean, µ. This is called **Inferential Statistics** and it is our final unit of this course. We will learn to estimate a parameter by building confidence intervals and performing hypothesis tests. This week, we will build confidence intervals about p, the true proportion.

A confidence interval will be stated in parentheses and listed: (lower bound, upper bound).
The width of the interval depends on the level of confidence. As the confidence level increases, the lower bound gets lower and the upper bound gets higher.

The most common confidence levels are 90%, 95% and 99%, although you can build a confidence interval with any level of confidence. As you increase the level of confidence, you build a wider interval. A wider interval is not more accurate. It is just more likely to capture the true proportion since you cast a "bigger net". We need to build confidence intervals that are accurate, not just wide. To do this, you want to make the standard error small and this is accomplished by making n large (increasing the sample size).

A confidence interval is calculated as follows: Point Estimate $\pm$ Margin of Error.

Recall that p is the symbol for the population proportion (true proportion) and $\hat{p}$ is the symbol for the sample proportion. In this unit, we are interested in estimating the population proportion, p, by using the best estimate we have. Our best estimate for p is called the point estimate for p and it is the sample proportion $\hat{p}$. To build a confidence interval, we will build an interval around the sample proportion by adding and subtracting a 'margin of error'.

Margin of Error (ME) is the maximum expected difference between the population proportion and the sample proportion. To calculate the margin of error, we multiply the standard error (SE) by the number of standard errors away we want to go from the center (Z_{crit}). The Z_{crit} is a Z-value, a location, for the number of standard errors away from the sample proportion. Margin of Error is often denoted ME.

$ME = (Z_{crit})\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$

The margin of error is both added to and subtracted from $\hat{p}$. The result is a lower bound and upper bound.

Our best estimate for p is the sample proportion. We say that our Point Estimate for p is $\hat{p}$.

$$\hat{p} \pm (Z_{crit})\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

The lower bound (LB) is found by taking $\hat{p} - ME$.

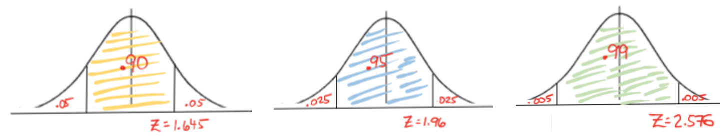
The upper bound (LB) is found by taking $\hat{p} + ME$.

The standard error will be calculated using the following formula. Note, we just used this standard error formula in describing the spread of a sampling distribution of $\hat{p}$. In this lesson, we do not know p, the true proportion, so we have to use $\hat{p}$ in the formula for standard error:

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

A 95% confidence interval means that if we were to take 100 different samples and compute a 95% confidence interval for each sample, then approximately 95 of the 100 confidence intervals will contain the true proportion (**p**). In practice, however, we select one random sample and generate one confidence interval, which may or may not contain the true proportion. The observed interval may over- or underestimate **p**. Consequently, the 95% CI is the likely range of the true, unknown parameter, p.

If you were building the CI by hand, you would need to choose the Z critical value specifically for the level of confidence. Here are some commonly used values for 90%, 95% and 99%. Recall that we reviewed three important Z-critical values in the lesson from last week.



For a 90% Confidence Interval, use: $\hat{p} \pm (1.645)\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

For a 95% Confidence Interval, use: $\hat{p} \pm (1.96)\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

For a 99% Confidence Interval, use: $\hat{p} \pm (2.576)\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

If we used the same sampling method to select 100 random samples and build 100 different confidence intervals, we would expect the population proportion to fall within 95 of the confidence intervals. In other words, a 95% level of confidence indicates that we expect 95% of the intervals built to capture the population proportion.

We will use technology to build confidence intervals (CI). In my video introduction below, **I build a CI by hand** to give some background knowledge.

Interpreting a Confidence Interval (CI)

I suggest using one of the following interpretation templates:

"We are ___% confident that the true (population) proportion _________ is between LB and UB."
Describe the variable here

OR

"We estimate, with ___% confidence, that the true (population) proportion _______is between LB and UB."
Describe the variable here

We have confidence that the population proportion is in the interval we built. We don't know for certain. When we build a 95% confidence interval, the we say we are 95% confident that the true proportion is in the interval.

Saying population proportion is the same as saying true proportion. This is also the same as saying the proportion of all ______. You can use any of the phrases to describe the proportion of the population.

Meeting the Assumptions/Requirements for a Z-Curve

There is one requirement that needs to be met before we can trust the results of a CI built using a Z-curve: The Normality Assumption. Based on our work with sampling distributions, we will use the following calculation to determine if the sample size is considered large enough to assume a normal distribution: $(n)(\hat{p})(1-\hat{p}) \geq 10$. NOTE: We do not know p, so we must use $\hat{p}$ as an estimate for p.

THE NORMALITY ASSUMPTION MUST BE MET IN ORDER TO TRUST THE CI RESULTS FROM THE Z-CURVE.

I am fully aware that there are other ways to check if the normality assumption has been met. I am asking you to use this method: $(n)(\hat{p})(1-\hat{p}) \geq 10$.

Answering Follow-Up Questions about the CI

A confidence interval can be used to answer follow questions. Recall that the true proportion is a proportion in the interval, with some confidence. Any proportion in the interval could be the actual population proportion.

It is also important to note, that in order for the confidence interval to support a statement, the entire interval must support the statement.

Majority

NOTE: **Majority** means more than 50%. In order for a CI to support that a majority feel a certain way, the entire CI would need to be greater than .50.

NOTE: We will be using technology to build confidence intervals in this course. If you use StatHelper, the results show the work in building the lower and upper bounds of the confidence interval. Minitab 19 only produces the results. The StatHelper and Minitab 19 keystrokes are on the next pages:

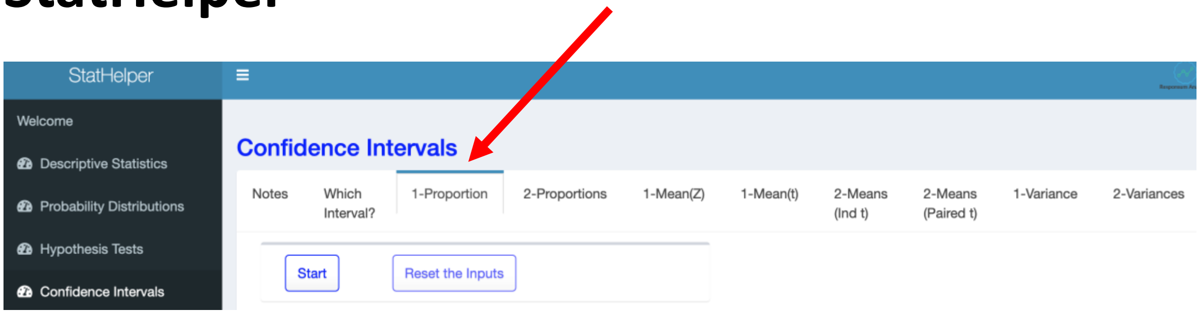
We will use technology to build confidence intervals (CI). In my video introduction below, **I build a CI by hand** to give some background knowledge.

Video Introducing the CI formula for p with Example from Prof. Coffey

Here is a video discussing Intro to CI: https://youtu.be/hsqA61-JkZw

Here is the same video with interpreting: https://youtu.be/e4XxMWqBwKY

Building a Confidence Interval about p –with StatHelper



StatHelper Instructions

1. Choose Confidence Intervals from the left side.
2. Choose **1-Proportion** from the options across the top.
3. Press Start.
4. **Input the Confidence Level:** 90, 95, 99, etc.
5. **Select the Type of Interval**: 2-Sided Interval.
6. **Input the Number of Events**: This is X, the number of responses that you received.
7. **Input the Number of Trials (n)**: This is n, the total number that were sampled.
8. Click RUN. Your confidence interval is under both the WORK tab and the INTERPRETATION tab.

NOTE: Some statisticians, including the creator of StatHelper, use π to represent the population proportion and p to represent the sample proportion. We will use p to represent the population proportion and p-hat or $\hat{p}$ to represent the sample proportion.

Example 2 with StatHelper

A researcher is interested in the proportion of all x-ray machines that malfunction and cause excess radiation. A random sample of 140 machines is taken and 42 of the machines are found to malfunction.

A. Estimate the proportion of all x-ray machines that malfunction, with 95% confidence. In other words, build a 95% confidence interval for the **true** proportion of x-ray machines that malfunction. In other words, build a 95% confidence interval for the **population** proportion of x-ray machines that malfunction.

> Think about it:
> Before you get started, be sure you can identify the variable and describe the population.
> The variable is whether or not the x-ray machine malfunctions; it is CATEGORICAL (we counted those that malfunctioned).
> The population is ALL x-ray machines.
> $\hat{p} = \frac{42}{140} = .30$ The number of events = 42; The number of trials = 140

StatHelper Output:

> **Now we can combine the LB and UB to get the 95% confidence interval:**
>
> $\pi \in (0.2241, 0.3759)$

INTERPRETATION: We are 95% confident that the true proportion of x-ray machines that malfunction is between .22 and .38.

ASSUMPTION MET? Yes. Calculate: (140)(.3)(1 - .30) = 29.4 Since the product is at least 10, we know that the sample size is large enough to assume a normal distribution.

B. Your colleague wants to use this headline in your research paper: "Less than 25% of x-ray machines malfunction!" Can we support this statement, with 95% confidence? Explain.

No, we are unable to support this statement with 95% confidence. In order to support this statement, the entire CI would need to be less than .25 and it is not; only some of the interval is less than .25. We do not know the true proportion of defective x-ray machines but, with 95% confidence, we believe it is between 22% and 38%. We cannot support this statement.

C. Your boss is interested in using this statement: "Twenty five percent of x-ray machines malfunction!" Can we support this statement, with 95% confidence? Explain.

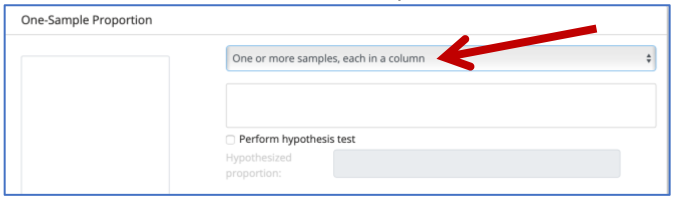
Yes, we can since .25 is in the confidence interval. Any proportion in the interval could be the true proportion so saying .25 is the population proportion is reasonable, with 95% confidence.

D. If, instead, we were to build a 90% confidence interval, explain how this would affect the confidence interval.

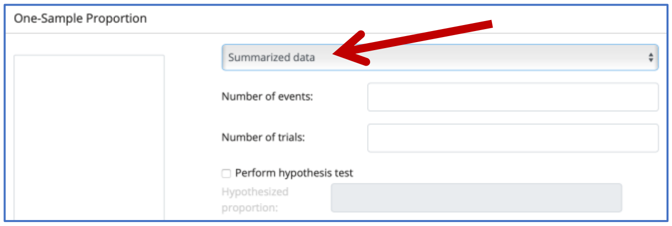
When you decrease the level of confidence, the Zcrit value decreases and, therefore, the margin of error decreases and the interval becomes narrower.

Building a Confidence Interval about p –with Minitab

<u>Minitab 19 Keystrokes</u>
Stat → Basic Statistics →1 Proportion

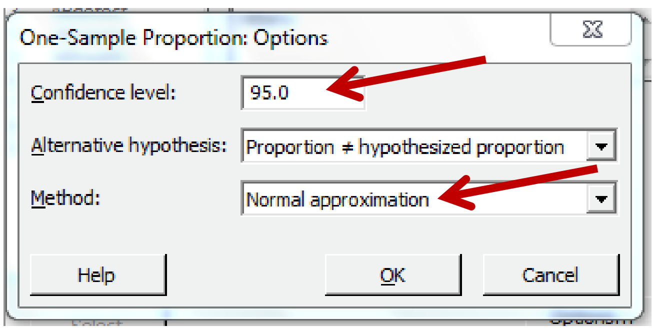


Choose 'One or more samples, each in a column' if you have raw data; i.e. your categorical data is in a column of Minitab. We will not work with raw data in this lesson.



Choose 'Summarized Data' if you do not have raw data. The data was already counted and you were provided the number of events (X) and the number of trials (n).

Recall: $\hat{p} = \frac{x}{n}$



Click on OPTIONS
Adjust the confidence level (default 95)
Alternative should show ≠(default)
Change Method: NORMAL APPROXIMATION

Stat→Basic Statistics→1 Proportion

1. Bring in your data/statistics

- If you have raw data in a column in Minitab, you will bring that column in.
- If you have summarized data, change the drop-down arrow to 'Summarized Data'.

2. Select '**Options'** and change the confidence level as needed.
3. Change the Method to read 'Normal Approximation'.

Example 3 with Minitab 19

The city council in a large city has become concerned about the trend toward exclusion of renters with children in apartments within the city. To determine the extent of this problem, they selected a random sample of apartments and determined for each whether children are permitted or excluded. It was found that 37 of the 250 sampled apartments <u>exclude</u> renters with children.

A. Estimate the population proportion of apartments in the city that exclude renters with children with 90% confidence.

> Think about it:
> Before you get started, be sure you can identify the variable and describe the population.
> The variable is whether or not apartments exclude renters with children; It is CATEGORICAL (we counted those that were excluded).
> The population is ALL apartments that rent to those with children in this city.
> $\hat{p} = \frac{37}{250} = .148$ The number of events = 37; The number of trials = 250

Minitab 19 Output

WORKSHEET 1

Test and CI for One Proportion

Method

p: event proportion
Normal approximation method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	90% CI for p
250	37	0.148000	(0.111059, 0.184941)

INTERPRETATION: We are 90% confident that the true proportion of apartments that exclude renters with children is between .11 and .18.

ASSUMPTION MET? Yes. Calculate: (250)(.148)(1 - .148) = 31.5 Since the product is at least 10, we know that the sample size is large enough to assume a normal distribution.

B. The housing coordinator has said in a recent meeting that more than 10% of apartments prohibit renters with children. Does the CI support the housing coordinator's statement, with 90% confidence? EXPLAIN.

Yes, with 90% confidence, we can say 'more than 10%' since the entire CI is greater than 0.10.

C. If we wanted to have 95% confidence (instead of 90% confidence), explain how this would affect the confidence interval.

When you change the level of confidence to 95%, the Zcrit increases and, therefore the margin of error increases and the interval would be wider.

D. If the sample had n = 500 (with p-hat still equal to 0.148), explain how this would affect the confidence interval.

If the sample size increases, the standard error decreases. If the standard error decreases, then the margin of error decreases and the interval would be narrower.

More Examples with Follow-up Questions

Example 4 (Practice with follow-up questions)

In the Weymouth, MA health survey there were 333 adult respondents who reported a history of Type II diabetes out of 3573 respondents (333/3573=0.0932 or 9.32%). A 95% confidence interval is calculated to be (0.084, 0.103) or 8.4% to 10.3%.

A. Interpret the interval in the context of the problem.

We are 95% confident that the population proportion of adults in Weymouth, MA with a history of Type II diabetes is between 8.4% and 10.3%.

B. Check to be sure the normality assumption has been met.

Calculate: (3573)(.0932)(1 - .0932) = 301.97. Since the product is at least 10, we know that the sample size is large enough to assume a normal distribution. Yes, the assumption has been met.

C. Local health officials believe that the rate of Type II diabetes in Weymouth, MA adults is the same as the US rate of 9.4%. With 95% confidence, does the CI calculated above support this statement?

Yes, it is reasonable to assume the rate is 9.4% since this value is in the 95% confidence interval.
(NOTE: Each value in the confidence interval is a reasonable rate.

D. A recent article states that more than 8% of adults in Weymouth have Type II diabetes. Does the CI above support this statement, with 95% confidence?

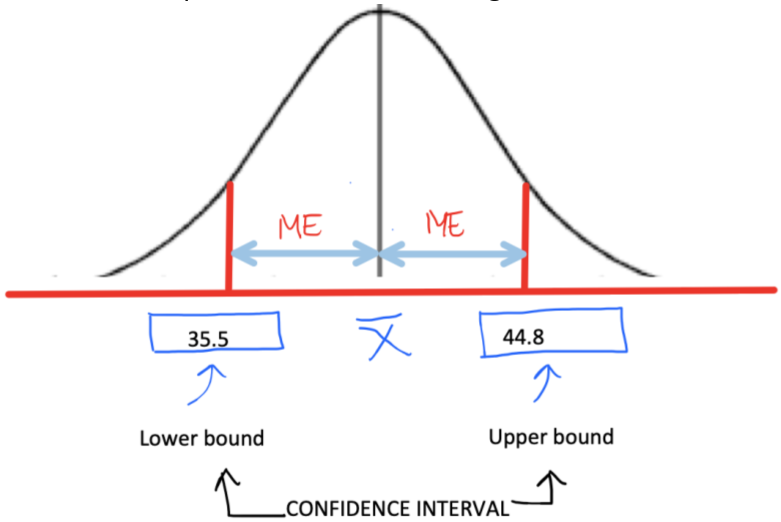
Yes, the entire CI is greater than 8%, therefore the CI supports the statement that the rate is more than 8%.
NOTE: For a CI to support a greater than or less than statement, the entire CI must support the statement.

E. Another article states that more than 9% of adults in Weymouth have type II diabetes. Does the CI above support this statement, with 95% confidence?

No, with 95% confidence, the CI does not support this statement since some of the values in the CI are less than 9%. For the CI to support the statement, all the values in the CI must support the statement.

Working with an Existing CI

If a confidence interval is provided to you as (lower bound, upper bound), you can work backwards to determine the point estimate and margin of error.



The **point estimate** is the value in the center of the CI. To find it, take the average of the lower and upper bounds.
The **margin of error (ME)** is the distance from the point estimate to either the upper or lower bound. For example, I can take the upper bound minus the point estimate to get the margin of error (ME) or I can take the point estimate minus the lower bound to get the ME. Alternatively, it is the difference between the upper and lower bounds divided by 2.

Example 5 (Find Point Estimate and ME)

We are 95% confident that the true proportion of RIT students that smoke is between .175 and .236.
Using your knowledge of confidence intervals, find the point estimate and margin of error.

ANSWER: The point estimate is the value in the center of the CI. To find it, take the average of the lower and upper bounds. 0.2055.

The margin of error (ME) is the distance from the point estimate to either the upper or lower bound. I will take the upper bound minus the point estimate to get the margin of error (ME). 0.236 - 0.2055 = 0.0305.

Alternatively, I can find the margin of error by finding the difference from the lower bound to the upper bound and dividing by 2. (.236 -.175)/2 = .061/2 = .0305

Example 6 (Majority)

We are 95% confident that the true proportion of first year RIT families that attend Brick City Homecoming on any given year is between .42 and .53.

Using the confidence interval, are we able to say that a majority of first-year RIT families attend Brick-City Homecoming?

No, majority means ‘more than 50%’ and since the entire confidence interval is not greater than .50, I cannot say that a majority attend.

Understanding How a CI Changes

Let's look again at the anatomy of the CI formula and discuss what happens if certain values were to be changed.

$$\hat{p} \pm (\boldsymbol{Z_{crit}})\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

QUESTIONS YOU NEED TO BE ABLE TO ANSWER:

What happens to the CI if we decrease the level of confidence to 90%?

A 90% CI is a narrower interval compared to a 95% CI. The change in the formula would be as follows: The Z-value associated with 90% confidence is smaller than for 95% confidence, which makes the calculation of the margin of error smaller which results in a narrower CI.

What happens to the CI if we increase the level of confidence to 99%?

A 99% CI is wider than a 95% CI. The change in the formula would be as follows: The Z-value associated with 99% confidence is bigger than for 95% confidence, which makes the calculation of the margin of error bigger, which results in a wider CI.

What happens if we increase the sample, but all other values remain the same?

If n increases, then the overall calculation of the standard error $\left(SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$ would decrease. Notice that n is in the denominator of the standard error. As the denominator gets bigger, the overall value of SE gets smaller. If the standard error (SE) decreases, then the overall margin of error decreases which results in a narrower CI.

What happens if the standard deviation of the sample increases, but all other values remain the same?

If s increases, then the standard error (SE) also increases. If SE increases, then the ME increases and the interval is wider.

Video Discussing the CI Formula from Prof. Coffey

Here is a video discussing the CI formula: https://youtu.be/edbZP-RVglM

Here is the same video with interpreting: https://youtu.be/k-c1dDY8-ec

Example 7 (Selecting a CI)

I have calculated a 90% confidence interval as (0.56, 0.63). Which of the following could possibly be a **95%** confidence interval for the same sample?

A. (0.547, 0.617)

B. **(0.547, 0.643)**

C. (0.573, 0.643)

A 95% confidence interval would be wider. That means that the lower bound would get smaller and the upper bound would get bigger. This only happens in B.

Estimating Sample Size with Categorical Data

We are going to work with the part of the CI formula that represents margin of error:

$$\hat{p} \pm \underbrace{(Z_{CRIT})\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{\text{This is Margin of Error (ME)}}$$

$$ME = (Z_{CRIT})\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

A researcher may have an idea about the margin of error she is comfortable with. She may also know what level of confidence she is interested in. Sometimes, a previous research study can provide the researcher with a decent estimate (prior estimate) for the value of p-hat. With all of this information, the researcher can then determine the sample size needed to achieve these results. **Always round UP to the nearest integer.**

If we use algebra to solve the above equation for n, we end up with the following:

$$n = \hat{p}\,(1-\hat{p})\left(\frac{Z_{CRIT}}{ME}\right)^2$$

This formula can be used to solve for sample size (n) if

- p-hat can be estimated with a prior estimate. If we are not given a prior estimate, just use a conservative estimate p-hat = 0.50.
- Z_{CRIT} can be determined if you know the level of confidence (i.e. if 90% confidence, use Z_{CRIT} = 1.645, if 95% confidence, use Z_{CRIT} = 1.96, if 99% confidence, use Z_{CRIT} = 2.576)
- ME or margin of error is given in the problem; If given as a percent, convert to decimal.

Video Estimating Sample Size with Example from Prof. Coffey

Here is a video estimating sample size: https://youtu.be/hWMksPR-bO8

Here is the same video with interpreting: https://youtu.be/u5-FOdpjkwU

Example 8 (Estimate Sample Size)

Last year a study showed that 17% of RIT students smoke cigarettes. You are interested in conducting research to estimate the true proportion of RIT students that smoke cigarettes with 99% confidence and you are comfortable with 3% margin of error. Find the sample size needed to achieve these results.

- n = ?
- p-hat from prior study = .17
- Z_{CRIT} = 2.576 (The ZCRIT for 99% confidence)
- ME = 3% = .03

$$n = \ .17\,(1-.17)\left(\frac{2.576}{.03}\right)^2$$

$$n = \ (.1411)(85.8666667)^2$$

$$n = \ (.1411)(7373.08444)$$

$$n = \ 1040.342215$$

$$\textit{Always round UP to the nearest integer}: n = \ 1041$$

ANSWER: We should sample 1041 RIT students.

Example 9 (Estimate Sample Size)

To estimate the next president's final approval rating, how many registered US voters should be sampled so that the approval rating is within 2.5 percentage points, with 95% confidence? Use President Bush's final approval rating of 0.22 as an educated guess. Always round UP to the nearest integer.

- n = ?
- p-hat from prior study = .22
- Z_{CRIT} = 1.96 (The ZCRIT for 95% confidence)
- ME = 2.5% = .025

$n = (.22)(1-.22)\left(\frac{1.96}{.025}\right)^2$ = 1054.75

ANSWER: We should sample 1055 registered US voters.

Example 10 (Estimate Sample Size with NO PRIOR ESTIMATE)

Insurance companies are interested in knowing the population percent of drivers who always buckle up before riding in a car. When designing a study to determine this population proportion, what is the minimum number you would need to survey to be 90% confident that the population proportion is estimated to within 0.02?

- n = ?
- p-hat from prior study = UNKNOWN…so we will use .50
- Z_{CRIT} = 1.645 (The ZCRIT for 95% confidence)
- ME = .02

$n = (.50)(1-.50)\left(\frac{1.645}{.02}\right)^2$ = 1691.3

ANSWER: We should survey 1692 drivers. **<u>NOTE: This example is completed in the video above.</u>**