Can you guess what we are gonna study???

Unit 7 – Confidence Interval Estimates

Your Confident Professor Colton

Unit 7 - Outline

Unit 7 – Confidence Interval Estimates

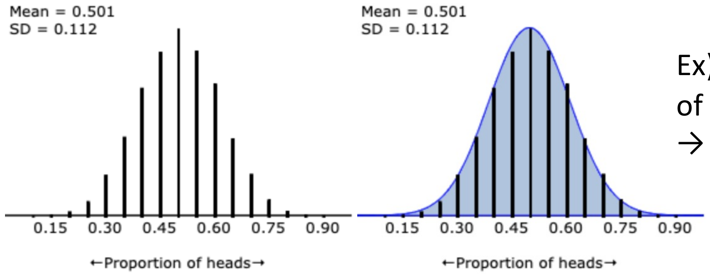
Sampling Distribution and CLT of $\hat{p}$ Review

Intro
- Populations and Samples

Proportions

- Motivating Example
- How to build a CI
- Margin of Error
- Formula
- Interpreting CI
- Practice

Central Limit Theorem for $\hat{p}$



Ex) Results from 10,000 sampl
of 20 coin tosses
→ *p = 0.5*

REVIEW!

<u>Central Limit Theorem</u>

- Let $\hat{p}$ be the sample proportion of successes in a random sample of size n from a population with true proportion of success p.
- If we take a <u>large enough sample</u>, then
  - The mean of $\hat{p}$ is equal to the population proportion, *p*

$$\mu_{\hat{p}} = p$$

($\mu_{\hat{p}}$ in words = mean (center) of distribution of sample proportions)

($\hat{p}$ = "p-hat" = sample proportion)

  - The standard deviation of $\hat{p}$ is equal to

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

($\sigma_{\hat{p}}$ in words = Standard deviation of sample proportions)

** Might see $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$, where q = 1- p
(it represents the probability of failure)

  - And the distribution of $\hat{p}$ is approximately Normal!

*** Again, technically there are conditions for this (same idea, but different because proportions now). But we will ignore them ☺*

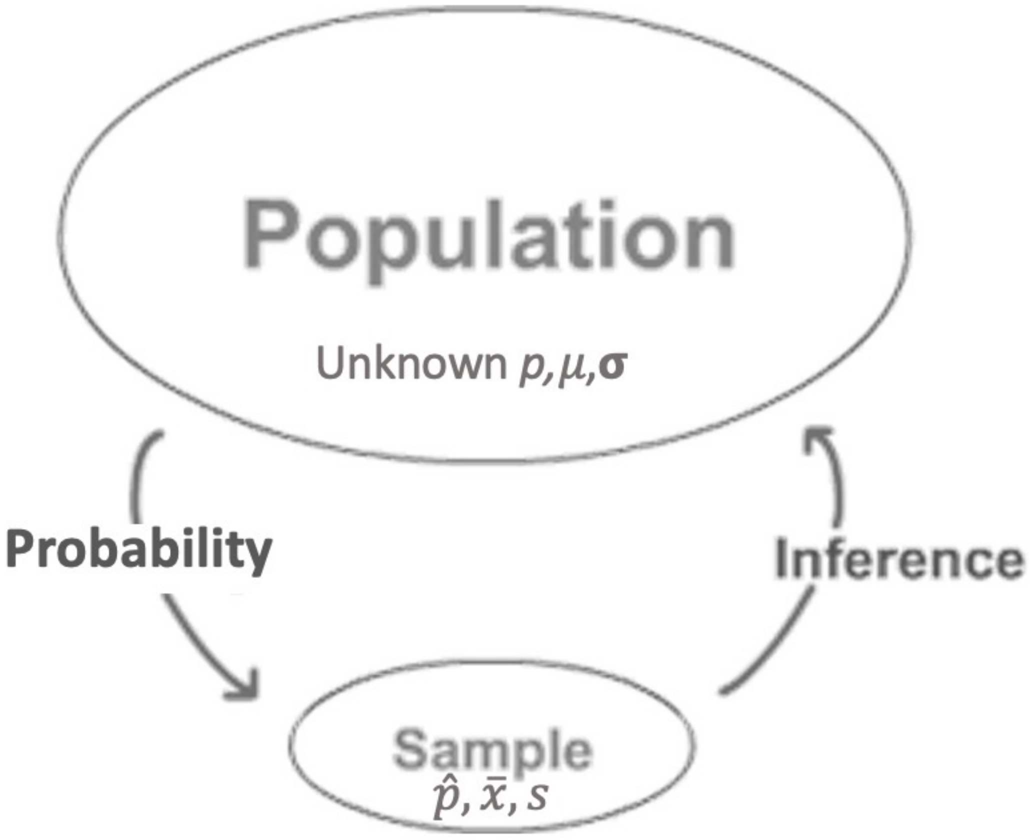
$$\hat{p} \sim Normal\left(mean = p, SD = \sqrt{\frac{p(1-p)}{n}}\right)$$

<u>*Summary*</u>

$\hat{p}$ is Normal with mean $\mu_{\hat{p}} = p$ and SD $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

- *(Referring to the sampling distribution, selecting a group of people and summarizing)*

Motivation – Populations and Samples



Motivating Example

Last time taught these slides, I lost

Future teaching strategy:
I think slides are quality and there s

Just don't over explain, they know v

The more I explain before they have

Setup: A random sample of 100 fourth-graders were surveyed to determine if they a

- The sample proportion for the 100 students was $\hat{p} = 0.45$
- Administrators want to estimate what the proportion would be if the entire population of fourth graders in the district had been surveyed.

Building up:

- The best estimate for the unknown population proportion is the sample proportion $\hat{p} = 0.45$, which is a point estimate
- Let's say we think that $\hat{p}$ could be off by 5% from the unknown population proportion, we would estimate p with the interval $0.45 - 0.05 < p < 0.45 + 0.05 \ \rightarrow \ 0.45 \pm 0.05$

And we're there:

- The interval (0.4, 0.5) is called a confidence interval and stats peeps construct confidence intervals to estimate unknown population parameters!
  - *Construct confidence intervals means finding the lower bound and upper bound*
- The plus-or-minus number is called the margin of error!

Motivation – Why Confidence Intervals

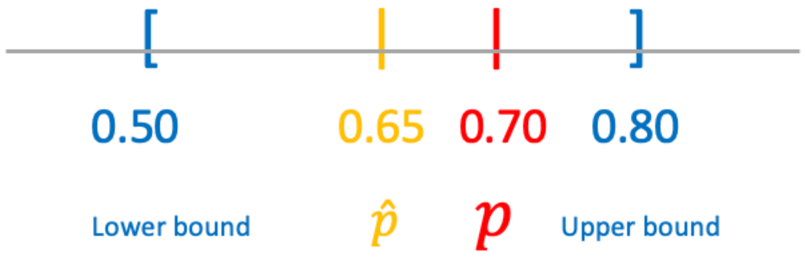
Estimating Parameters

Point Estimates

- Using a statistic to estimate a parameter (for means we use $\hat{p}$ or $\bar{x}$ to estimate p *or* μ, respectively)
- It is a single number that is our best guess (estimate).
- Very unlikely that statistics equal the true parameter values they are estimating (remember each sample is different; sampling variability).
- Therefore, in order for the estimate to be useful, we must describe how close it is likely to be.

Interval Estimates

- Give a range for what we think the population parameter is.
- Takes into account sampling variability.

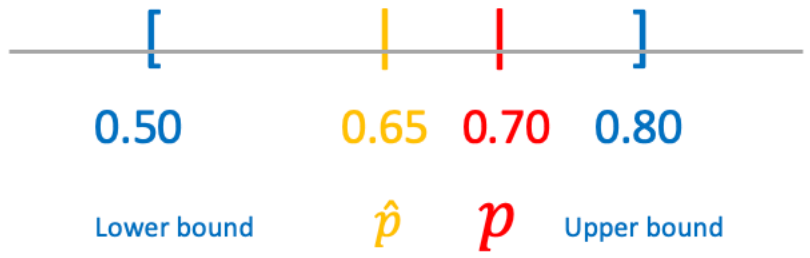


Now we are ready to dive deeper!

How do we formally build this interval?

What are the different pieces that make up a confidence interval?

How do we interpret the final interval? What does it mean?



Return to Motivating Example

How large to make the margin of error?

- We need to determine how large to make the margin of error so that the interval is likely to contain the population proportion (this is the GOAL, to capture the parameter).
- To do this, we use the sampling distribution of $\hat{p}$, specifically the CLT formulas for the standard deviation of p-hat $\sigma_{\hat{p}}$
  - For this example, the sample size is n = 100 and the proportion that we know is $\hat{p}$ = 0.45. So the standard error (think standard deviation) is $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.45\,(1-0.45)}{100}} \approx 0.0497$
- There is one more piece to the margin of error…

Confidence Level and Critical Value

- Every confidence interval must have a confidence level.
- The confidence level is a percentage between 0% and 100% that measures the success rate of the method used to construct the confidence interval.
- How to we incorporate this into our interval calculation??

How to Build a Confidence Interval

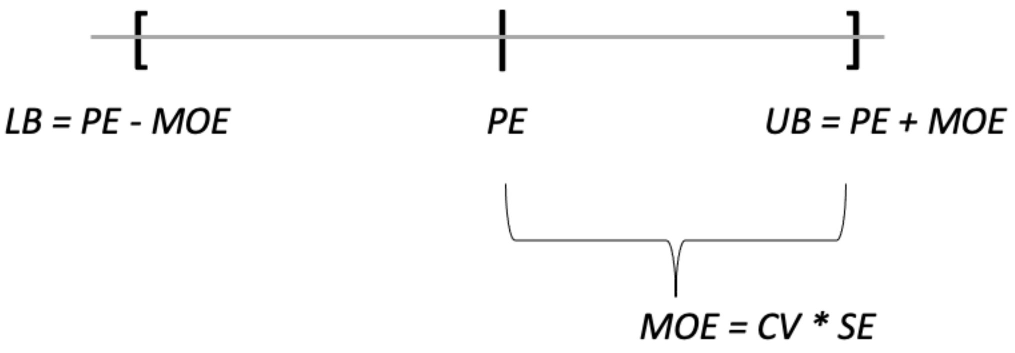
C.I. = Point Estimate ± Margin of Error

Point Estimate

- **Point Estimate** is your best guess; at the center of the interval.
  - Then we extend our guess in both directions in order to provide a wider range of plausible values.
  - This distance is called the **Margin of Error**.

Margin of Error

- **Margin of Error (MOE)** is what makes our estimates intervals rather than just single points!
  - Made up of two components that will be discussed on the next slide!
  - MOE = Critical Value (CV) * Standard Error (SE).



Margin of Error

Margin of Error

- This determines how much wiggle room we have around our sample proportion
- MOE = Critical Value (CV) x Standard Error (SE)

  $= Z^* \, \sigma_{\hat{p}}$

  $= Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

MOE = (how many steps to take) x (how big is each step)
*** this is another way to think about it*

Now let's breakdown the two pieces!

Standard Error

- Measures sampling error.
- The *standard deviation of the sampling distribution* $\hat{p}$ is

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

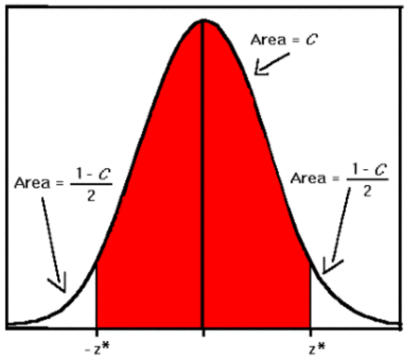
- When we don't know p (i.e. when making CIs), we substitute $\hat{p}$ in for p and it becomes

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

- This quantity is now referred to as the *standard error*!

Critical Value – Z Star

- In a CI, the point estimate (center) and the SE (step size) are both based on the collected data.
  - There is nothing we can do as the researcher after the fact to change these…
- But we CAN control the Confidence Level of our interval ('We are 95% Confident' or 90%, etc)!
  - *We do this by changing the Critical Value (CV), Z**
  - Said another way, we decide how many steps to take in each direction from our sample proportion!
- The specific values for Z* are just the Z-scores that mark the middle %C of the standard normal curve!
- We are going to be given these values in a table!



Confidence Level (%C)	Critical Value Z*
80%	1.28
85%	1.44
90%	1.64
95%	1.96
98%	2.33
99%	2.58

Return to Motivating Example

<u>Construct final interval: C.I. = Point Estimate ± Margin of Error</u>

Margin of Error

- In our example, the standard error was $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.45\,(1-0.45)}{100}} \approx 0.0497$ and the critical value for a 95% confidence level is 1.96.
- So Margin of Error = 1.96 x 0.0497 ≈ 0.097

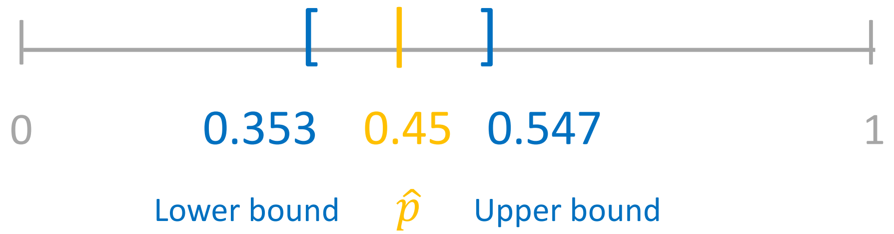
= 1.96 steps of size 0.0497

If we wanted to be more confident, take more steps

Confidence Interval

- Point Estimate – Margin of Error < p < Point Estimate + Margin of Error
- The <u>point estimate</u> was $\hat{p} = 0.45$ and <u>MOE</u> = 0.097. So a 95% confidence interval is:

$$\hat{p} - 0.097 < p < \hat{p} + 0.097$$
$$0.45 - 0.097 < p < 0.45 + 0.097$$
$$\mathbf{0.353 < p < 0.547}$$



Final Confidence Interval for *p*

1 Proportion Z Interval

Recall: Our point estimate is the sample proportion $\hat{p} = \frac{x}{n}$, which represents the number of success divided by the sample size.

C.I. = Point Estimate ± Margin of Error

$$= \hat{p} \pm Z^* \sigma_{\hat{p}}$$

$$= \hat{p} \pm Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow \left(\hat{p} - Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Summarizing LCQ!

Setup

A NatGeo Poll interviewed 1200 hiking enthusiasts and asked "Are you more afraid of spiders or snakes???" Out of the 1200 people, 768 responded "Ewww, snakes….". **Calculate** the corresponding *95% confidence interval*!

Solution

- p = ??
- $\hat{p} = ??$
- CV = ??
- $SE = ??$
- $MOE = ??$
- $95\%\ CI = ??$

Summarizing LCQ!

Setup

A NatGeo Poll interviewed 1200 hiking enthusiasts and asked “Are you more afraid of spiders or snakes???” Out of the 1200 people, 768 responded “Ewww, snakes….”. **Calculate** the corresponding *95% confidence interval*!

Solution

- *p* = *What is the context?? In words, p represents the true proportion of hikers that are more afraid of snakes*

- $\hat{p} = \frac{x}{n} = \frac{768}{1200} = 0.64$

- *CV* = Z^* = *1.96*

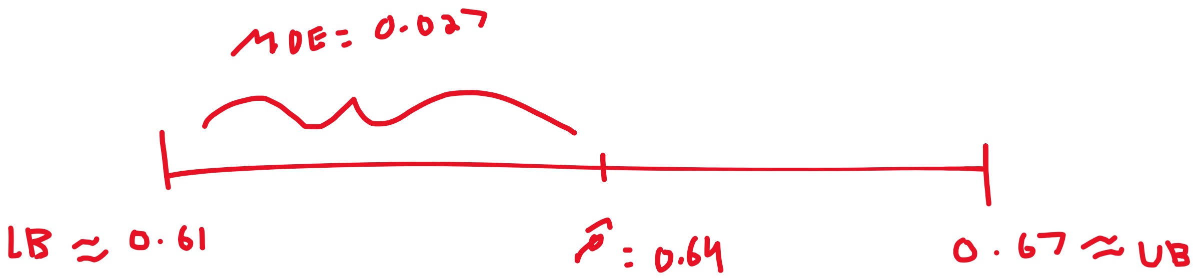
- $SE = \sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.64\,(1-0.64)}{1200}} = 0.0139$

- $MOE = Z^*\sigma_{\hat{p}} = 1.96 * 0.0139 = 0.0272$

- $95\%\ CI = \hat{p} \pm MOE = 0.064 \pm 0.0272 = (0.6128,\ 0.6672)$

*** This is how we would have to do it by hand!*

But we are going to use the Calculator! Phew!



Using Calc!

<u>Setup</u>

A NatGeo Poll interviewed 1200 hiking enthusiasts and asked "Are you more afraid of spiders or snakes???" Out of the 1200 people, 768 responded "Ewww, snakes….". **Calculate** and **interpret** the corresponding *95% confidence interval*!

<u>GOAL</u>: Find the Confidence Interval!

1. 1-PropZInt
   a) x = # of successes (people that said yes)
   b) n = sample size
   c) C-Level = Confidence level (as a decimal or whole number, both work)

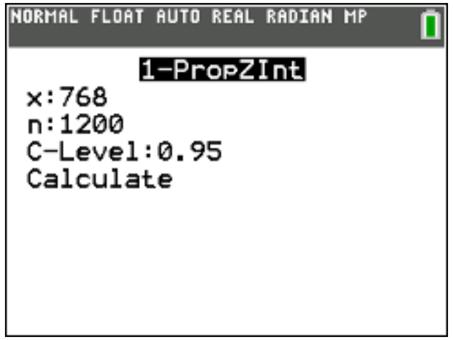
<u>Interpret results</u>:
??

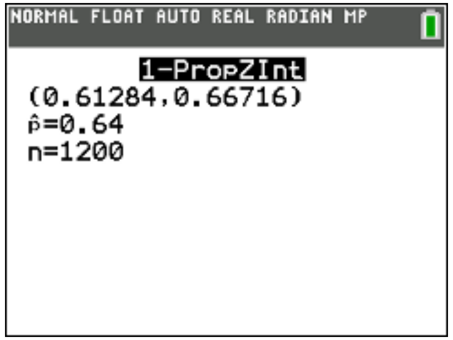
Using Calc!

<u>Setup</u>

A NatGeo Poll interviewed 1200 hiking enthusiasts and asked “Are you more afraid of spiders or snakes???” Out of the 1200 people, 768 responded “Ewww, snakes….”. **Calculate** and **interpret** the corresponding *95% confidence interval*!

<u>GOAL</u>: Find the Confidence Interval!





1. 1-PropZInt
   a) x = # of successes (people that said yes)
   b) n = sample size
   c) C-Level = Confidence level (as a decimal or whole number, both work)

Show work: 95% CI = 1-PropZInt(x = 768, n = 1200, C-Level = 0.95) → (0.6128, 0.66716)

*** Can just copy and paste the general structure and fill in the information and results for this specific problem!!*

I am <u>% confident</u> that the true/population <u>parameter + context</u> is between <u>(lower bound)</u> and <u>(upper bound)</u>.

<u>Interpret results</u>:

I am <u>95% confident</u> that the true <u>proportion of hikers who are more afraid of snakes than spiders</u> is between <u>0.6128</u> and <u>0.6672</u>.

Interpreting Confidence Intervals

General Structure

I am C% confident that the true/population parameter + context is between (lower bound) and (upper bound).

True = population (they mean the same thi
Parameter will either be MEAN or PPROPO

Example

Trying to estimate the proportion of all Columbus residents who enjoy running → 95% CI = (0.05, 0.25)

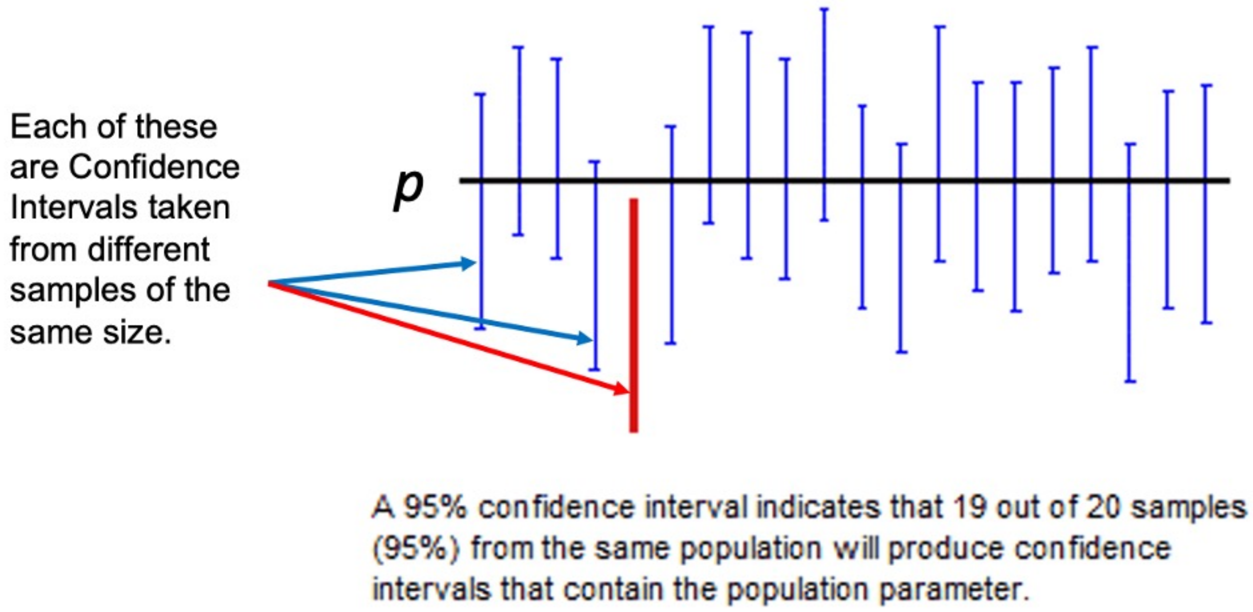
- We are 95% confident that the true (population) proportion of all Columbus residents who enjoy running is between 0.05 and 0.25.

3 Pieces

1. 95% Confident: This is a Confidence Statement
   - Tells us what percent off ALL possible samples result in a CI that captures the true proportion.
2. Parameter + Context: We are talking about population proportions.
   - But what population proportion??? We ALWAYS need context.
3. Interval: The range of plausible values!
   - Uses our sample statistic and MOE.

Interpreting Confidence Intervals

Confidence Interval Interpretation Visualized



Dope applet!

Very Important!

- The confidence level is NOT the probability the parameter is in the interval.
- It refers to the long run capture rate (i.e. over many, many intervals constructed in the same way).
- Either the interval contains the parameter or it does not.

Another LCQ

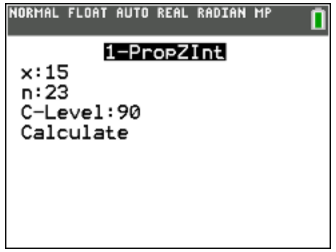
Setup: 15 out of 23 people from a random sample said their National Championship team is still remaining in their NCAA March Madness Bracket.

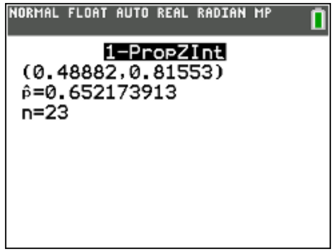
1) Calculate the 90% Confidence Interval.

2) Interpret this interval.

Another LCQ

Setup: 15 out of 23 people from a random sample said their National Championship team is still remaining in their NCAA March Madness Bracket.





1) Calculate the 90% Confidence Interval.

90% CI = 1-PropZInt(x = 15, n = 23, C-Level = 90) → (0.489, 0.816)

Writing out the calculator function and inputs like this is how we would show work for calculating our interval!

2) Interpret this interval.

We are 90% confident that the true proportion of people who's national championship team is still remaining in their March Madness bracket is between 0.489 and 0.816. → This a PERFECT interpretation!

Examples of INCORRECT (and common) interpretations!

There's a 90% chance that the sample proportion of people who's national championship team is still remaining is between 0.489 and 0.816. → TWO things wrong

- *'confident' means something specific in Statistics, do NOT want to use the word 'chance' 'probability', etc.*
- *We are trying to estimate the TRUE or the POPULATION proportion, that's the goal. NOT the SAMPLE proportion, we already know what that is. So do NOT say 'sample'*

We are 90% Confident that the population proportion is between 0.489 and 0.816. → MISSING CONTEXT! Have to say what this proportion represents!

Another LCQ

Problem: From a random sample of 65 students, 40% said they prefer to wake up early to do their homework rather than stay up late.

Calculate and interpret the 98% confidence interval.

Solution

Another LCQ

Problem: From a random sample of 65 students, 40% said they prefer to wake up early to do their homework rather than stay up late.

Calculate and interpret the 98% confidence interval.

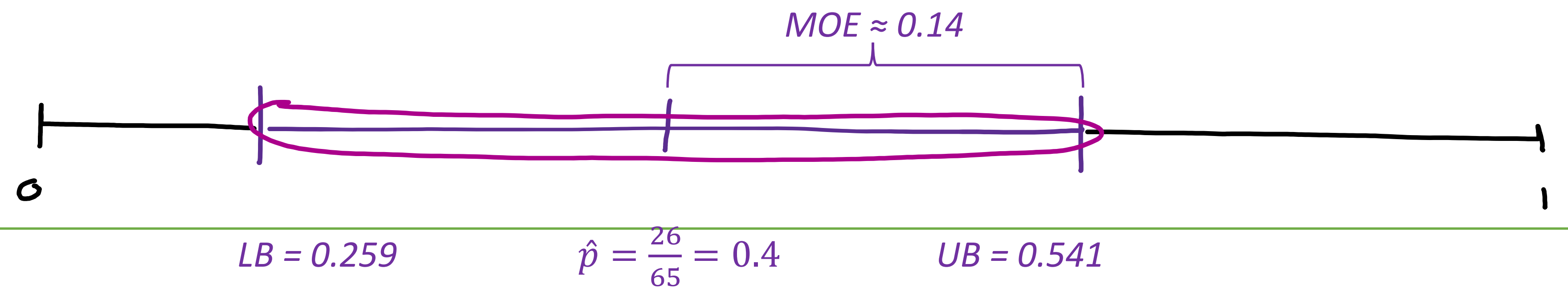
Solution

Calculate Interval:

- *1-PropZInt*
  - *x = 26 (=0.4(65))*
  - *n = 65*
  - *Confidence = 0.98*
- *Result = (0.25864, 0.54136)*

Interpret Interval:

We are 98% confident that the true proportion of students who prefer to wake up early to do their homework is between 0.259 and 0.541



One more LCQ

Setup: From a random sample 500 people, 64% said they prefer to vacation at the beach compared to the mountains.

Recall: Our point estimate is the sample proportion $\hat{p} = \frac{x}{n}$, which represents the number of success divided by the sample size.

1) Calculate the 85% Confidence Interval.

2) If I increase the sample size to 600 (*and keep* $\hat{p}$ *= 0.64 → new x = 600(0.64) = 384)*, what will happen to the new confidence interval (wider, narrower, stay the same)?

3) If I change the Interval from Question 1 to be 90% Confident, what will happen to the new confidence interval (wider, narrower, stay the same)?

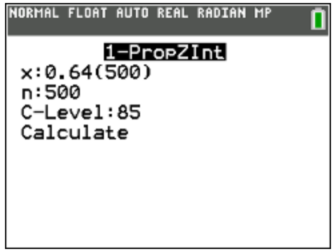
One more LCQ

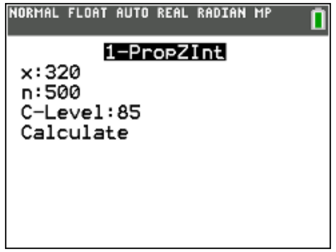
Setup: From a random sample 500 people, 64% said they prefer to vacation at the beach compared to the mountains.

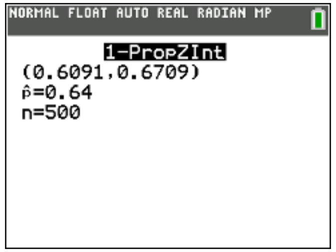
1) Calculate the 85% Confidence Interval.

85% CI = 1-PropZInt(x = 320, n = 500, C-Level = 0.85) → (0.6091, 0.6709)

- *Have to type in x, but weren't given it directly*
- *So need to calculate it using $\hat{p}$ and n*







2) If I increase the sample size to 600 (*and keep $\hat{p}$ = 0.64 → new x = 600(0.64) = 384)*, what will happen to the new confidence interval (wider, narrower, stay the same)?

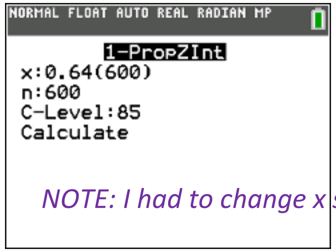
Interval becomes narrower!

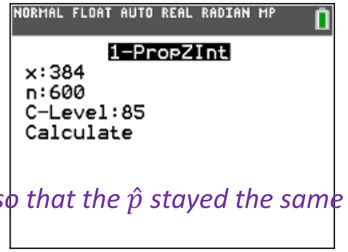
- *This is because of the MOE, and specifically the standard error!*

MOE = $Z^*\sigma_{\hat{p}}$

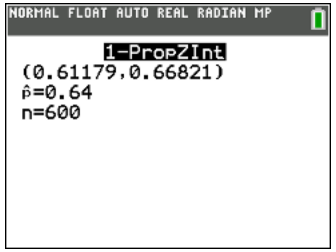
$$= Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- *If the confidence level stays the same, then the Z* value doesn't change*
- *BUT with a larger sample size, in the standard error equation we are dividing by a larger number.*
- *Thus making that overall quantity smaller and the MOE smaller → and confidence interval becomes narrower!*





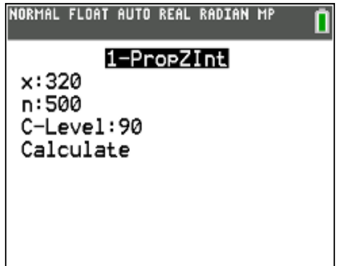
NOTE: I had to change x so that the $\hat{p}$ stayed the same

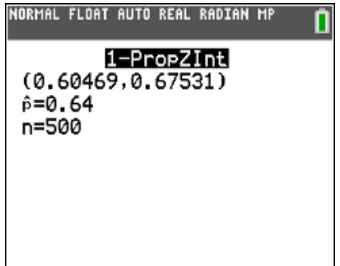


3) If I change the Interval from Question 1 to be 90% Confident, what will happen to the new confidence interval (wider, narrower, stay the same)?

It becomes wider.

- *Same idea as number 2, BUT now the standard error remains the same (because of the same n and $\hat{p}$).*
- *And the Z* value changes because of the new confidence level!*
- *To be more confident, we need to cover more values.*
  - *So have a larger critical value*
- *This makes the MOE increase and our CI become wider!*





Summary of ideas from previous LCQ

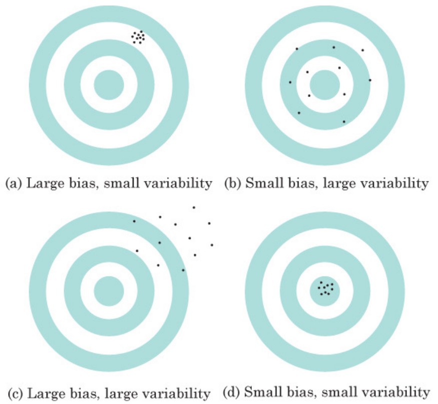


Figure 3.3
Moore/Notz, *Statistics: Concepts and Controversies*, 9e, © 2017 W. H. Freeman and Company

Precision

- Thinking back to sampling variability in surveys from Unit 1, we wanted to have good sampling methods so that our sample statistics could be precise (precision is GOOD)!
- This same concept is present with Confidence Intervals!

Precision with Confidence Intervals

- There are two ways to get a **more precise (narrower) confidence interval**!

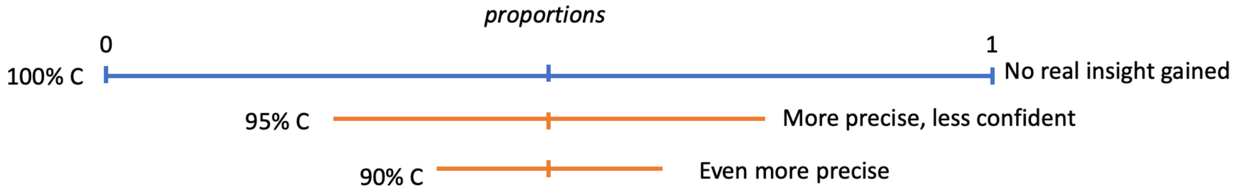
1. Increase the sample size (before collecting data GOOD!)
   - This decreases the standard error and as a result the MOE.

*(** Assuming everything else remains the same)*

2. Decrease the confidence level (after the fact, not so good)
   - This decreases the critical value and as a result the MOE.

Tradeoff between Precision and Confidence

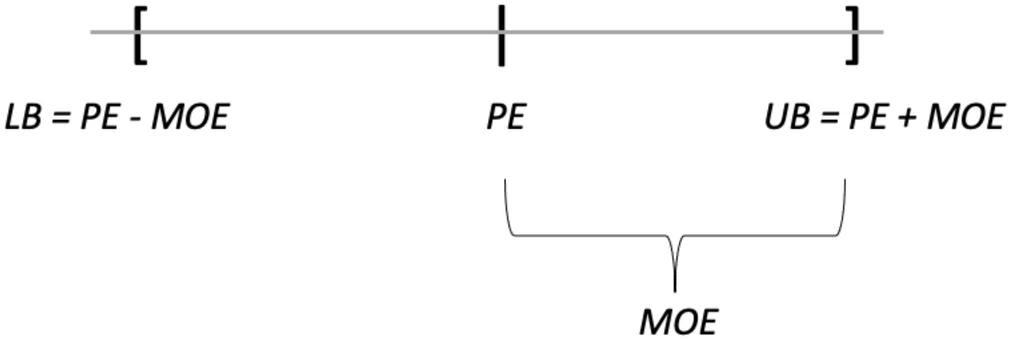
- If we want to be more confident, we need to cover more values!
- That of course decreases the precision…
- So there a pro / con to being super confident



Margin of Error Revisited

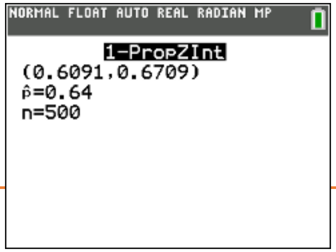
Recall: **MOE** is what you <u>add and subtract</u> from the <u>point estimate</u> to get the <u>bounds</u> of the <u>confidence interval</u>.

- If you are given an interval, we can start by finding the **width** of the interval:
    - Width = UB - LB
- Then we can easily find the **margin of error**:
    - Margin of Error = $\frac{Width}{2} = \frac{\text{UB}-\text{LB}}{2}$
- This gives us another expression for the **width:** Width = 2 * MOE



<u>Example</u>:

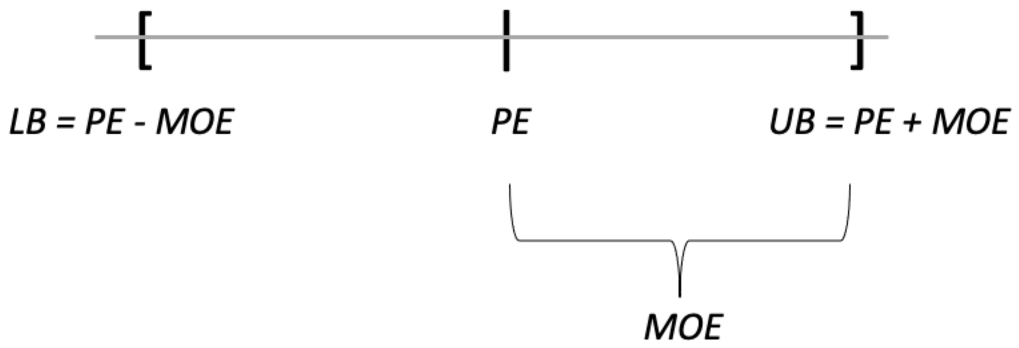
Find the Width and MOE based on the following output:



Margin of Error Revisited

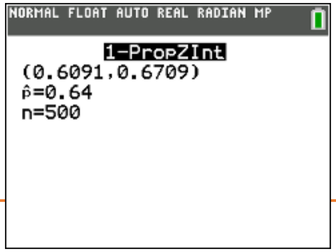
Recall: **MOE** is what you <u>add and subtract</u> from the <u>point estimate</u> to get the <u>bounds</u> of the <u>confidence interval</u>.

- If you are given an interval, we can start by finding the **width** of the interval:
  - Width = UB - LB
- Then we can easily find the **margin of error**:
  - Margin of Error = $\frac{Width}{2} = \frac{\text{UB}-\text{LB}}{2}$
- This gives us another expression for the **width:** Width = 2 * MOE



<u>Example</u>:

Find the Width and MOE based on the following output:



Width = UB – LB = 0.6709 – 0.6091 = 0.0618
MOE = width / 2 = 0.0618 /2 = 0.0309