STA-1380 – Elementary Statistics
Week #8

Hello! Welcome to the additional online **Weekly Resources** for the course of **STA-1380**. Following a traditional calendar semester, these will be some of the topics your professors will go over. If you do not see material your section is going over for the week, please look at the other resources listed for this course. In addition to these resources, there might be **Group Tutoring** for this course, please see our website for more details. These sessions will go over these materials in more detail as well as any questions about the material.

Any additional help or services can be found through the **Baylor Tutoring Website**. Visit to schedule a free 30-minute private tutoring session, drop-in times for your course, the Baylor Tutoring YouTube channel, or any additional tutoring resources.

Contacts: Sid Rich M-Th 9am-8pm (Fall and Spring class days) Office Phone: 254-710-4135

Topic of the Week:
"Estimating Population Parameters"

Key Points:

- Confidence Intervals
- Point Estimators

Gathering samples and their data collected can only be interpreted in so many ways. We are taught to assume that the $\mu_{\hat{p}} = p$ and $\mu_{\bar{x}} = \mu$, but what if our samples contained multiple outliers that skewed the results and makes our data remains unusable? To protect against these possibilities, statisticians use a device known as confidence interval to account for the possibilities of error while still using the data procured.

The subject of this week's guide will go over conceptual concepts such as what a Confidence Interval is and as to why we use them. Your teacher might not go over this section of the course in class, so this will be an additional tool should you need any further clarification. Remember that the goal of this is to use a sample and given knowledge of the study to provide an estimate for what the true population parameter should be.

Highlight #1
"Confidence Intervals"

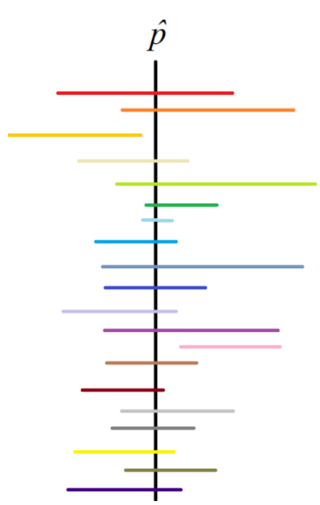
Confidence Intervals seek to provide a range of possible population parameter values. In other words, rather than using a single value, we will use a set area to which inferences about the population can be made. The term confidence comes from the percentages correlating to the correct value being chosen. A 90% confidence interval is formed using a function that produces a correct interval 90% of the time. Not that a single interval is right 90% of the time, but on average will be right 90% of the time.

The Confidence Interval DOES NOT mean that there is a 90% chance of the interval being correct or that there is a 90% probability that the interval is correct. One rule of thumb is that you must always 'tiptoe' around phrases that imply certainty, you cannot imply certainty in a Statistics course.

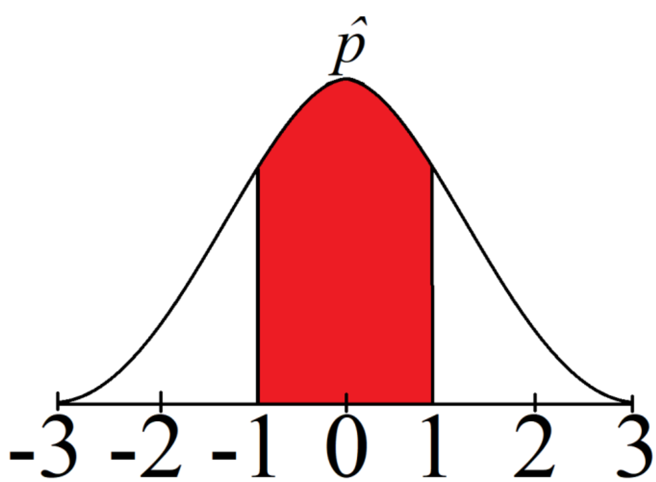
One way to correctly interpret a CI is by declaring your 'confidence' in the tools used, the other is to reference repeated trials. One way is to say that "We are 90% confident that true population parameter of X is found within the Interval of (Low, High)." Another is to say that "In repeated trials we would expect that 90% of the intervals contain the true population parameter of X and for it to be between Low and High." By doing this, you are not outright saying your choice is correct, but that you have reason to believe it to be, and you suggest others believe it as well.

When working with Confidence Intervals, it is important to remember that you are using data from samples, therefore, you will be using the $\hat{p}$ as both the point estimator and in making the Standard Error. Recall that the Standard Error is Std. Dev. over the square root of the sample size.

Confidence Intervals can be visualized with graphs, lines, and formulas:



*A line Chart demonstrating 90% Confidence Intervals



* A Confidence Interval on a Standardized Normal Distribution

The line charts are often used as a demonstration to show how in repeated samples, the confidence intervals capture the population parameter roughly as much as the confidence value. These are meant to show how varying Confidence intervals still manage to capture the population parameter regardless of width or size and how certain intervals may miss the mean despite being made the exact same way as the others.

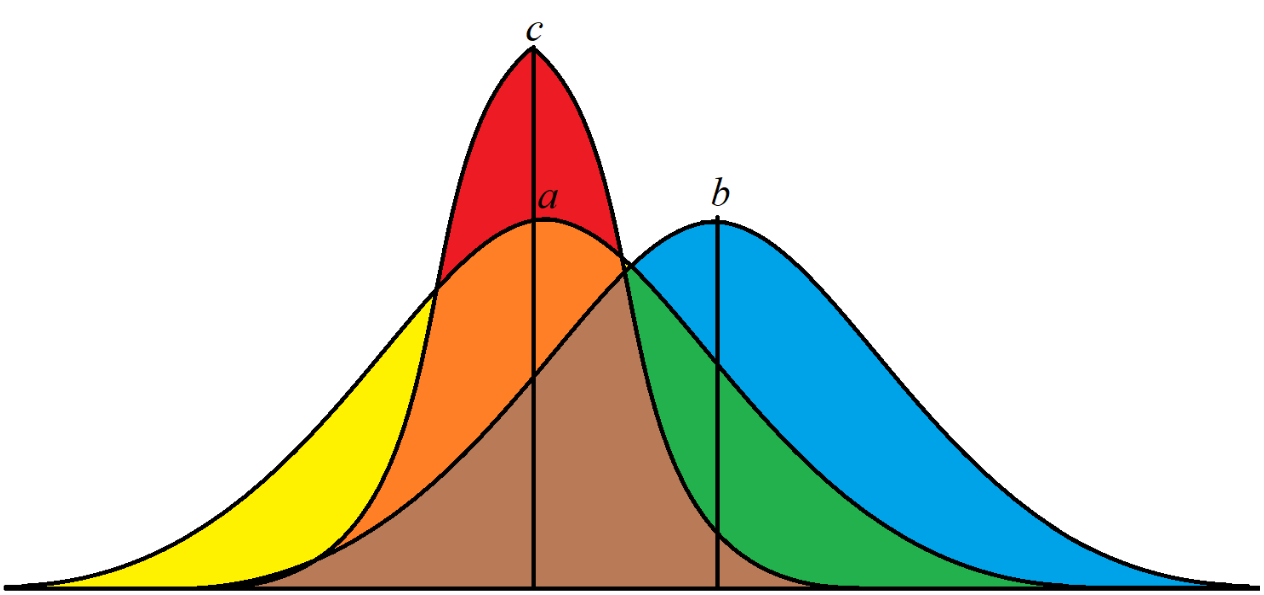
The graphs of a Confidence Interval will resemble that of an 'In-between' Normal Distribution, where the data you are looking for is in between two points, not including the tails of the graph. The area encompassing these graphs will be determined on the size of the interval. The larger the interval, the larger the range of data points. In other words, the more confident you wish to be in your interval, the larger the width of each side must be.

The formulas of a Confidence Interval are the nitty-gritty of this chapter and allow statisticians to form their inferences with real data points. These will be discussed in next week's guides.

Highlight #2
"Point Estimates"

Definition: A single value gained from a Sample or Sampling Distribution used to estimate the true value of its corresponding population parameter.

Example: $\mu_{\hat{p}}$ and $\mu_{\bar{x}}$ are both Unbiased Point Estimates for population proportions and means.
$\hat{p}$ and $\bar{x}$ are both Point Estimates for the population proportion and means.



*A figure depicting the differences between Bias and Variance between Estimators

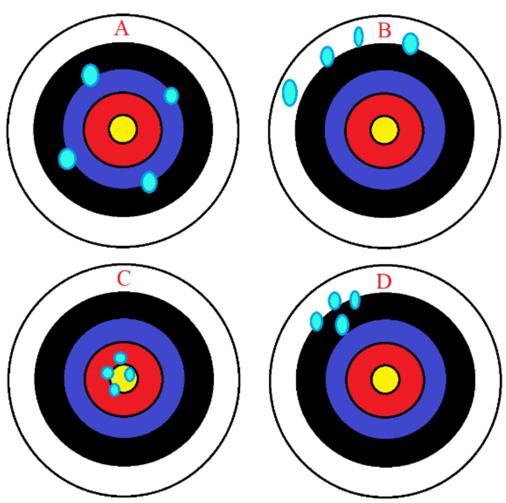
In last week's guide, we went over how the Sampling Distribution's center points, either for a mean or a proportion, was assumed to be equal to the population proportion. This is because the nature of a Sampling Distribution accounts for many trials that slowly but surely normalize the data. Those values are called Unbiased Point Estimators, which is a fancy term for saying that the values do a pretty good job at estimating what the true population values are. But what about regular $\hat{p}$ and $\overline{x}$? How can those two values be used to estimate the parameters of a population?

There are two things to look out for when choosing a point Estimator. Suppose that a scientist wishes to take a sample of Brazilian Tree Frogs to find what proportion of them are female. The Scientist has three potential values they can use as the point estimate for the population parameter decided (proportion of females).

When comparing value *a* with value *b,* the scientist realizes that they share the same spread of the data, but one accurately portrays the true population proportion, and the other is significantly off. The difference between *a* and *b* is that they share the same shape parameter, or that they have an equal variance, but one value, *b,* is a bias estimator. This means that it will not be able to accurately depict the true population because the center of the data is skewed. Between choices *a* and *b,* the Scientist should choose *a.*

When comparing value *a* with value *c*, the scientist realizes that they share the same location, but one has a significantly higher peak with smaller tails compared to the other. The difference between *a* and *c* is that they share the same location parameter, meaning that they are both unbiased estimators of the true population parameter, but one value, *a,* has significantly more variance. This means that it will not be able to accurately depict the true population proportion because more of the data will be spread out farther apart. Between choices *a* and *c*, the scientist should choose *c.*

Another way to think about this process is through a series of Arrow shots:



Target A represents an unbiased estimate with a high variance (Centered, Scattered)
Target B represents a biased estimate with a high variance (Not Centered, Scattered)
Target C represents an unbiased estimate with a low variance (Centered, Clumped)
Target D represents a biased estimate with a low variance (Not Centered, Clumped)

When given the choice between multiple potential point estimates, order them in terms of bias, then variance. Meaning that of the options, choose C first, A second, D third, and B fourth. You will use these Point Estimates as the center points of your confidence intervals.

Check Your Learning

1. In your own words, describe why a study would wish to use a current sample's mean as the estimate instead of a previously given standard?
2. What is one thing you should never say when declaring a Confidence Interval?

Things Students Struggle With

1. Confidence Intervals and their Distributions
   a. Because CIs are built around a sample we DO NOT USE A NORMAL DISTRIBUTION. It is vital for students to understand that we use a Sampling Distribution (either Normally distributed or a T-distribution you will learn in later chapters). Using a Normal Distribution is for a single instance, but because we have 'n' number of instances to account for, we use the sampling distributions from chapter 4 and use that as our bases. We never use a typical normal distribution, only ever a sampling distribution that is normally distributed.

2. Why?
   a. Confidence Intervals are unlike any other mathematical process done in this course before. Many students struggle to understand 'why' we do them. Why Estimate a range when we can use the sample mean on its own? The purpose is simple, we want to account for potential misdirection. A single number is good for estimating, but providing a range of points can be vital to understanding the number in context. The difference between a CI's range being + or – 3% versus 30% is massive depending on what requirements the estimate has to reach. This is why we use it, to give context to the people who will use the estimate after it has been calculated.

Concluding Comments

That's it for this week! Please reach out if you have any questions and don't forget to visit the Tutoring Center website for further information at https://www.baylor.edu/tutoring

Answers to CYL

1. A study wishes to estimate the current standard, using a current sample would be unbiased from the old standard and thus, a better estimate.

2. Never declare that the Confidence Level is the correctness of the Interval. Never say a there is a 90% chance that the CI is correct.