Week 9
Elementary Statistics: STA 1380

The concepts this resource covers are the topics typically covered during the first few weeks of the semester. If you do not see the topics your particular section of class is learning this week, please take a look at the weekly resources listed on our website for additional topics throughout the semester.

We also invite you to look at the group tutoring chart on our website to see if this course has a group tutoring session offered this semester.

If you have any questions about these study guides, group tutoring sessions, private thirty minute tutoring appointments, the Baylor Tutoring YouTube channel or any tutoring services we offer, please visit our website www.baylor.edu/tutoring or call our front desk during open business hours (M-Th 9 AM-8PM on class days) at 254-710-4135.

===

Week 9: Concepts

- SE, MOE, and Length
- CI Notation
- Estimating Sample Size

Short Review - Confidence

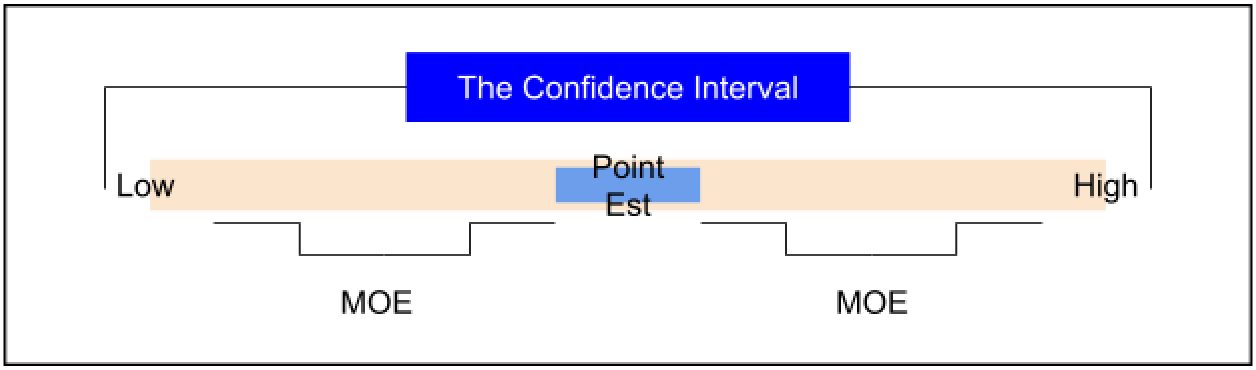
Before we approach the following topics, it would be important to consider what confidence intervals imply. They are a range of numbers with the center point being a point estimator. *(What is a point estimator?)* These bounds around the point estimator are the space by which the true parameter for the data exists. Remember to recall what the Z distribution implies and what the significance level α means. It would also be recommended to be familiar with the notation symbols like σ or μ to better read the questions and understand where some of the concepts and equations come from.

Concept #1 : SE, MOE, Length

Formulas: $SE = \frac{\sigma}{\sqrt{n}}$ or $\sqrt{\frac{\hat{p} \bullet \hat{q}}{n}}$ $\quad MOE = (Z^*_{\alpha/2})(\frac{\sigma_{\hat{p}}}{\sqrt{n}})$

Above are the formulas for the Standard Error which is considered the standard deviations of a sampling distribution. It is used only with sample data, meaning that it is denoted by common English letters instead of Greek letters so it is **not a parameter, but a sample statistic**. The same can be considered for the Margin of Error which is the distance that each point in the confidence interval falls from the mean. With a Confidence interval of (1,3), the MOE is equal to

1 and the point estimator is 2. This is because the point estimator is the center of the confidence interval, a common point estimator to use is the sample mean, sample proportion, or other cases that we will cover. The length of the confidence interval is equal to 2*MOE which would be the total length of the interval.



Concept Check: The term Margin of error is denoted as 'E' in your textbooks as 'error' for short. The Margin of error is the numerical value that is either added or subtracted from the PE to create the upper and lower bounds of the Confidence Interval. This is created by multiplying the SE with a particularly chosen Z*. The Length is x2 the MOE. It is the entire size of the Confidence Interval and captures the area both to the left and right of the Point Estimate. Knowing this, what would you say has a larger impact on the accuracy of the confidence interval? The Point Estimator or the Sample Size? The bigger the sample size the more narrow that the confidence interval becomes. This means that with too large of a sample size the confidence interval low and high values will become more similar. Due to this it is important to have good sample values as otherwise the interpretations can be incorrect. *Review: what is the interpretation of a confidence interval and what is the interpretation of a point estimator.*

Concept #2 : CI Notation

A confidence interval is generally written in standard notation of (low,high). The given interval of (3,4) would have a low of three and high of 4. The mean is the middle of the interval which can be found by using the average of the two points in the interval. The distance or "margin of error" is the space between each 'side' of the confidence interval and the point estimator or mean. In our example of (3,4), the length of the interval is 1, the point estimator is 3.5, and the margin of error is 0.5.

Concept Check: How are confidence intervals determined?

As you can recall from before, a normal distribution can be used to find the MOE, but what does this actually mean? Why do confidence intervals have greater lengths when the confidence is increased? Think back to what you have learned about normal distributions and Z scores. Would the Z score of 1.645 or 1.9 have a greater chance of occurring? The answer would be 1.9 as when reading the cumulative probability 1.9 is further from zero in a one sided confidence. However for two sided confidence intervals behave the same way where the further an associated Z score is from zero the larger the confidence interval is. *What is this relationship to alpha? Well as alpha grows smaller does confidence increase or decrease?*

<u>Concept #3: Estimating Sample Size</u>

As you saw above there is a relationship to how confidence intervals are calculated and the amount of samples. Through algebraic manipulation we can instead solve for n, which is our sample size. The following formula is a result of this transformation.

$$n = \frac{Z^{*2}(\hat{p} \bullet \hat{q})}{\text{MOE}^2}$$

When we observe this formula look at how the MOE has been squared as well as the Z score. This is a result of how the algebra moves the variables around. If you are interested in how this formula actually moves above you can find it by using the Z score equation. However for the scope of this course, focus on the aspects that would impact the size of n itself. This being the size of the margin of error or the p or q values. Raising the p value and/or shrinking the MOE will create a larger sample size and by extension raise the result of other variables.

If the p value is unknown then we can assume it to be 0.5. This is not particularly because of any specific statistical outcome but more as a baseline assuming an experiment has a 50/50 outcome set. This is the same idea behind why alpha is commonly assumed or set to be 0.05.

<u>Concept Check: Larger sample sizes have the greatest impact out of all changes a researcher can make?</u>

It's common for researchers to want to refine their process before actually starting. This is to save grant money, shorten the process, or just simply make the project go by easier. There are very few things that researchers can impact directly, these mainly being the alpha, sometimes beta, and most importantly the number of samples or n. Due to the following formula above when a large n is chosen in research design the researcher can “p-hack” or force the p value to be small. In hypothesis testing this is a *terrible* thing as it allows for research to be “proven” right when in fact it is incorrect. You will see this more so in the following weeks as it is considered one of the main two errors that we see in statistics.

Overall the main focus of choosing a sample size is to do so *after* you have found a margin of error and confidence. The p value being assumed is a way to prevent p hacking and one of the main ways to combat this. When working with sample size in particular it is helpful to make a checklist of what you know you have, the most common thing students will miss is that the p value can be assumed to be 0.5, so make sure to always account for that if it is not given or if you cannot solve for it by using the sample data given.

Example Questions:

1. Martha wants to see what proportion of students at Baylor are their parents' only children. She conducts a sample spanning all grades and majors to find that of the 1000 students sampled, 337 students were raised as the 'only-child' and 663 were found to have siblings.

   a. Construct a 95% Confidence Interval for p.

   b. Interpret what this interval means in terms of the context of the problem.

2. A Professor is grading the exams of "Stats 101," an online test done to assess the statistical knowledge of their students before the course begins. The Professor has hundreds of exams to assess to determine what proportion of their class has a minimum score of 75% on the test, this is the minimum score considered to be passing. If over 80% of the class passes, the professor can begin the course on Chapter 2, saving over 5 days of class time. If the population proportion is under 80%, they will have to start at chapter 1. Rather than hand-grade every test, they collect a random sample of 50 tests and compute the sample proportion of passing grades to be 88%.

   a. Using a 90% Confidence Interval, will the professor have to start on Chapter 1 or Chapter 2?

   b. Using a 95% Confidence Interval, will the professor have to start on Chapter 1 or Chapter 2? What does this say about an increase in CI level compared to the result?