Week 8
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 8: Concepts

- Confidence Intervals
- Point Estimators

Short Review - Normal Distribution and Z scores

Before we can begin to discuss confidence and its results: consider the following:

What does Z of α=0.05 mean?

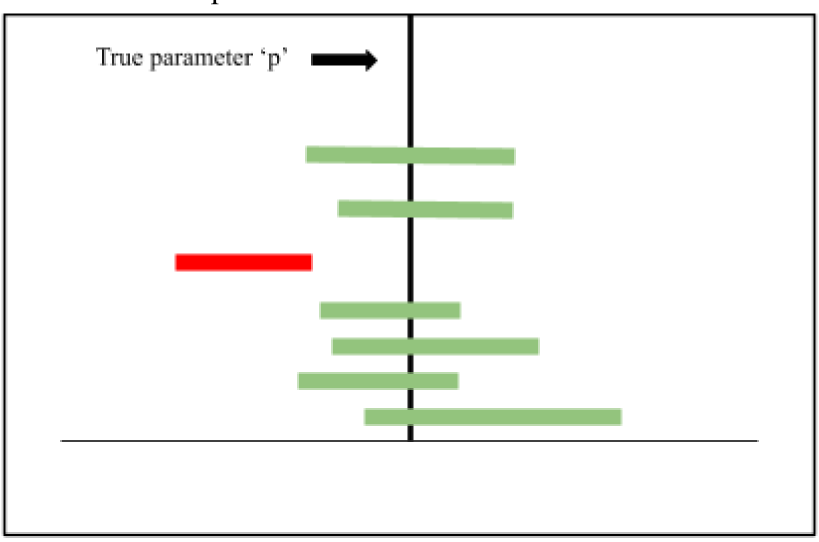
If you are unable to answer this question, please review earlier notes as the idea behind Z scores and the level of probability that they are associated with is critical to understanding the following concepts. If you were confidently able to say that this is the Z score that is responsible for accounting for 5% of the overall probability then you would be correct! But what does that mean for confidence? Think about Z scores in confidence as fence posts that we as statisticians aim for. It would be reasonable to assume that the mean of a sample of a population would be the same as the larger population but sometimes it's not, or our assumptions about the population were wrong. What if our samples contained multiple outliers that skewed the results and made our data remain unusable? To protect against these possibilities, statisticians use a device known as a confidence interval to account for the possibilities of error while still using the data procured.Hence we say as long as we are in the “fence-posts” of a confidence interval we have a sample that is representative of a larger population. *Take this with a grain of salt as later we will use this concept to try and identify if a sample behaves differently than its overarching population.*

Concept #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 doesn't mean that there is a 90% chance of the interval being correct. 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 the 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 p ˆas both the point estimator and in making the Standard Error. Recall that the Standard Error is sigma or your standard deviation over the square root of the sample size. Visuals of confidence intervals is as follows:



Notice how in the ranges of these confidence interval most of the time the true parameter is inside the range, but this does not always guarantee that its always is. Observe the red bar where the data was so extreme that the CI was pulled away from 'p'. This is what the interpretation of confidence means. If our confidence is 90%, then 90% of the intervals made from sample data would be green in this image, but 10% would be red and not have 'p' contained in the interval

Concept Check:Outliers

What is an outlier? A more formal definition is somewhat outside the scope of this course, but we can think of outliers as outside the fence post determined by the IQR or interquartile range. Some statisticians name them as points that are more than 3 standard deviations away from the mean, *why would this create extreme points knowing what you know about normal distributions?* The equation for the confidence interval follows as below:

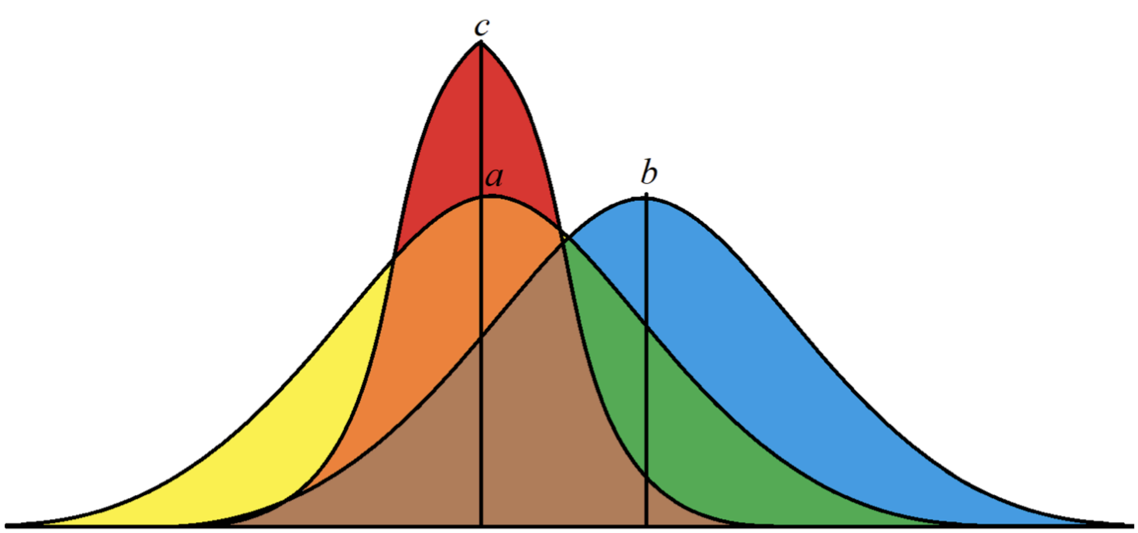
$$CI = \bar{x} \pm (Z * SE)$$

What this means is that depending on your mean (which can be heavily influenced by outliers) the entire interval can shift its location. Consider a set of numbers like 1,2,3 which has a mean of 2 vs a set of numbers from 1,2,9 which has a mean of 6. One extreme value can move x_bar so much that it will influence the entire position of a confidence interval.

Concept #2 : Point Estimators

Definition: A single value gained from a Sample or Sampling Distribution used to estimate the true value of its corresponding population parameter. Basically a statistic that we will use in order to make arguments about confidence.

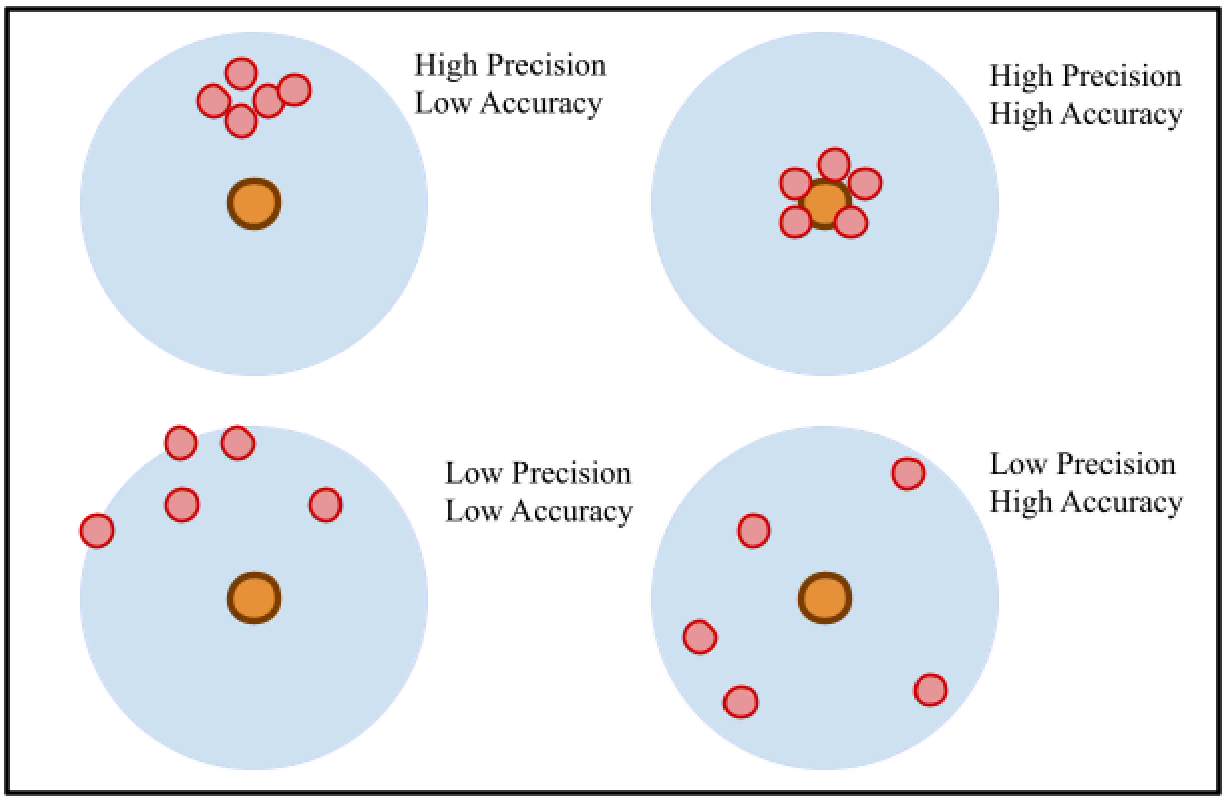
Example: $\mu_{\hat{p}}$ and $\mu_{\bar{x}}$ are both Unbiased Point Estimates for population proportions and means. $\hat{p}$ and $\bar{x}$ are both Point Estimates for the population proportion and means. The difference between an unbiased estimator and point estimator is the same reason that point estimators come from sample data.



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

Imagine from the following image above that A is the true distribution of the data. Where C is what happens when the data is very precise and or has a lower variance than the greater population. B occurs when the data is not as accurate compared to the true spread, but has been shifted most likely because of a series of outliers. When points and or overall point estimates are not good representatives of the larger data set we see changes like those above.

Use the following image to see the difference between Precision and Accuracy.



Here we see that the orange dots in the middle represent the means and the red dots represent samples in the blue circle sample spaces. When our data has low precision overall the confidence intervals and point estimates are the worst, but when accuracy is somewhat present these create the better quality confidence intervals. Accuracy is better than precision because of this.

Sampling Distributions account 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 $\bar{x}$? How can those two values be used to estimate the parameters of a population?

From the circles above:
Bottom Right represents an unbiased estimate with a high variance (Centered, Scattered)
Bottom Left represents a biased estimate with a high variance (Not Centered, Scattered)
Top Right represents an unbiased estimate with a low variance (Centered, Clumped)
Top Left 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 the highlighted options most frequently. Top Right is better than the Bottom Right but not all data behaves the way you'd want it to. You will use these Point Estimates as the center points of your confidence intervals.

Example Questions:
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?