**Confidence Interval**

From Sampling Distributions to Confidence Intervals

* We can use bootstrapping and sampling distributions to build confidence intervals for our parameters of interest.
* By finding the statistic that best estimates our parameter(s) of interest (say the sample mean to estimate the population mean or the difference in sample means to estimate the difference in population means), we can easily build confidence intervals for the parameter of interest.
* Imagine we have a sampling distribution of some statistic, mean, variance, etc.
* Can use this sampling distribution to build a confidence interval for our parameter of interest
* If we want a 95% confidence interval, we cut off 2.5% from both sides



* In this video, you saw an example of how to build a confidence interval using the sampling distribution of the statistic that best estimates your parameter of interest.
* In this case, we used a sample mean height to estimate the population mean height.
* You can interpret your confidence interval as **We are 95% confident, the population mean falls between the bounds that you find**.
* Notice that the percent and the parameter can both change depending on what you are building your confidence interval for, and what percentage you cutoff in each tail.

Difference In Means

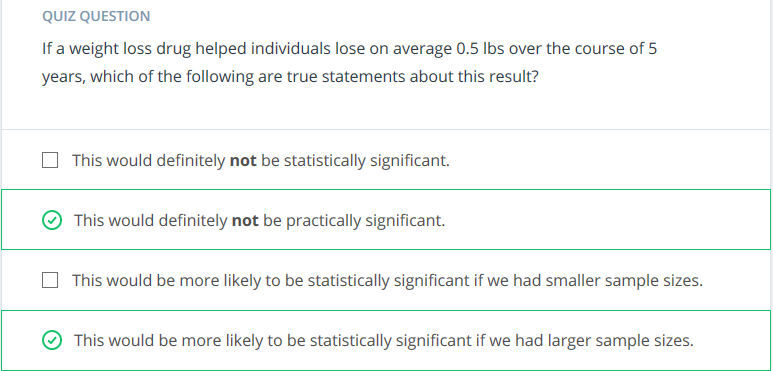
* In this video, you built a confidence interval for the difference of the average heights for coffee drinkers and non-coffee drinkers.
* The interval was built at a 95% confidence level, and since the difference did not contain zero, this suggested there was truly a difference in the average heights in the population of coffee drinkers as compared to non-coffee drinkers.
* Specifically, we can be 95% confident that the difference in the average heights for coffee drinkers as compared to non-coffee drinkers was in the provided interval of 0.59 to 2.37 inches.
* Notice the similarity of the wording to the last confidence interval you built. The highlighted portions signify the two parts that can change in your conclusions:
  + The confidence level.
  + The parameter you are capturing with your interval.

Confidence Interval Applications

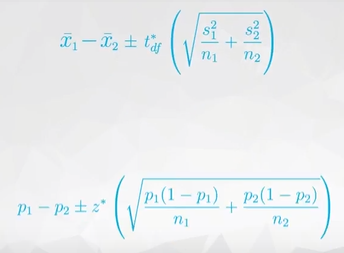
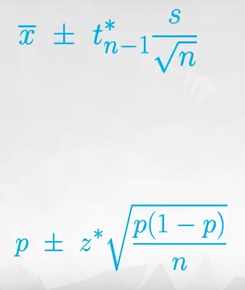
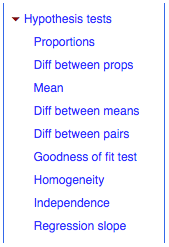
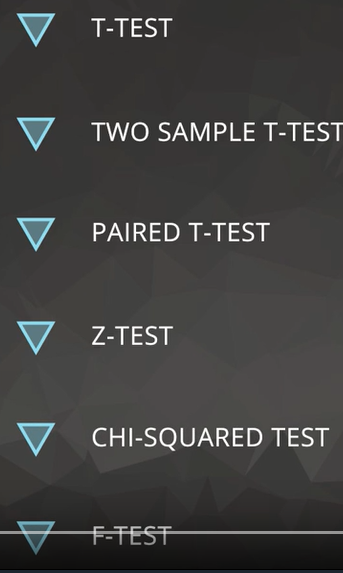
* Though you were comparing the average heights of coffee drinkers to non-coffee drinkers, there are a number of other applications that use a comparison for the means of two groups.
* A/B testing is one of the most important to businesses around the world.
* In this technique, you are changing something about your web layout to understand how it impacts users. You ideally want to provide a page that leads to more clicks, higher revenue, and/or higher customer satisfaction.

Statistical vs. Practical Significance

* Here, you learned about **practical** and **statistical** significance.
* Using confidence intervals and hypothesis testing, you are able to provide **statistical significance** in making decisions.
* However, it is also important to take into consideration **practical significance** in making decisions.
* **Practical significance** takes into consideration other factors of your situation that might not be considered directly in the results of your hypothesis test or confidence interval.
* Constraints like **space**, **time**, or **money** are important in business decisions. However, they might not be accounted for directly in a statistical test.



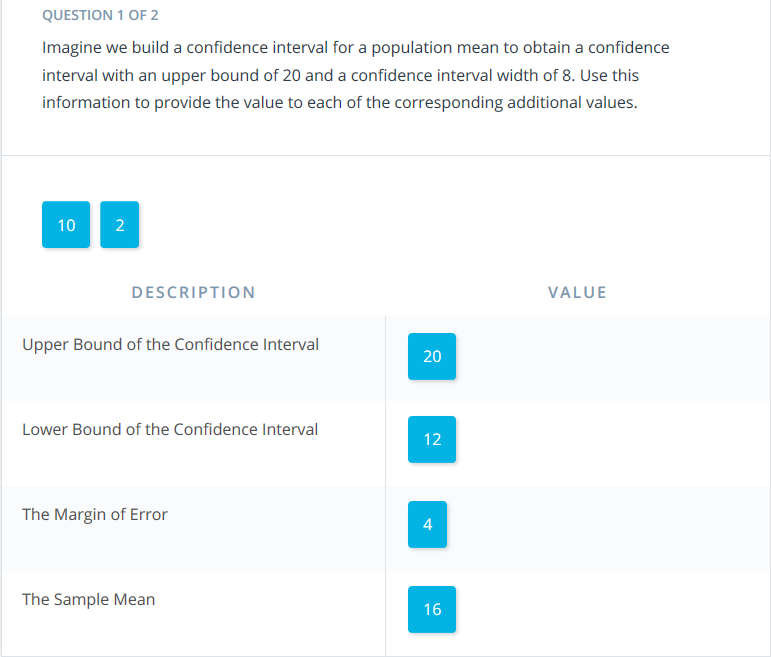
Traditional Confidence Intervals

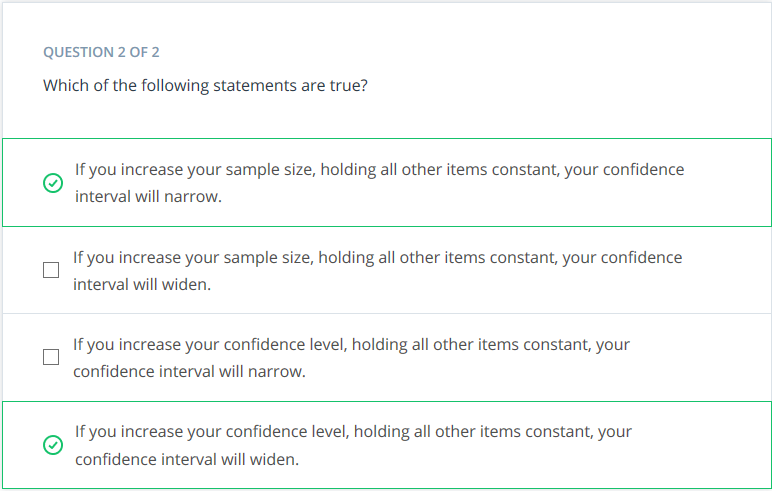
* Traditional confidence intervals methods have underlying assumptions about the data, which may or may not be true.
* Examples of traditional confidence intervals:
* 
* Bootstrapping on the other hand, does not have the same assumptions of these traditional confidence intervals.
* Bootstrapping only assumes the sample is representative of the population
* With **large** enough sample sizes, both methods provide essentially the **same** results
* With smaller sample sizes, using a traditional methods likely has assumptions that are not true of your interval.
* Small sample sizes are not ideal for bootstrapping methods though either, as they can lead to misleading results simply due to not accurately representing your entire population well.
* **One educated, but potentially biased opinion on the traditional methods** is that these methods are no longer necessary with what is possible with statistics with modern computing, and these methods will become even less important with the future of computing.
* Therefore, memorizing these formulas to throw at particular situation will be a glazed over component of this class.
* However, there are resources below should you want to dive into a few of the 100s if not 1000s of hypothesis tests that are possible with traditional techniques.
* To learn more about the traditional methods, see the documentation [**here**](http://stattrek.com/hypothesis-test/hypothesis-testing.aspx) on the corresponding hypothesis tests.
* In the left margin, you will see a drop down of the hypothesis tests available, as shown in the image below.
* 
* Each of these hypothesis tests is linked to a corresponding confidence interval, but again the bootstrapping approach can be used in place of any of these!
* Simply by understanding what you would like to estimate, and simulating the sampling distribution for the statistic that best estimates that value.
* 
* The above tests can be replaced with bootstrapping
* There are built-in packages in python that does bootstrap



Other Language Associated with Confidence Intervals

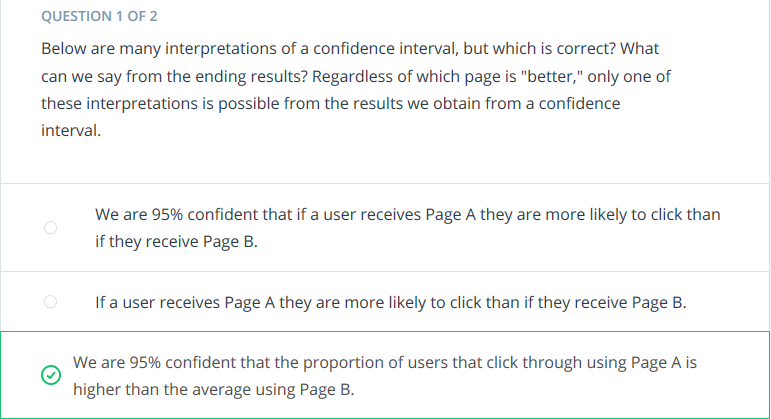
* It is important to understand the way that your sample size and confidence level relate to the confidence interval you achieve at the end of your analysis.
* Assuming you control all other items of your analysis:
  1. Increasing your sample size will decrease the width of your confidence interval.
  2. Increasing your confidence level (say 95% to 99%) will increase the width of your confidence interval.
* You saw that you can compute:
  1. The confidence interval **width** as the difference between your upper and lower bounds of your confidence interval.
  2. The **margin of error** is half the confidence interval width, and the value that you add and subtract from your sample estimate to achieve your confidence interval final results.

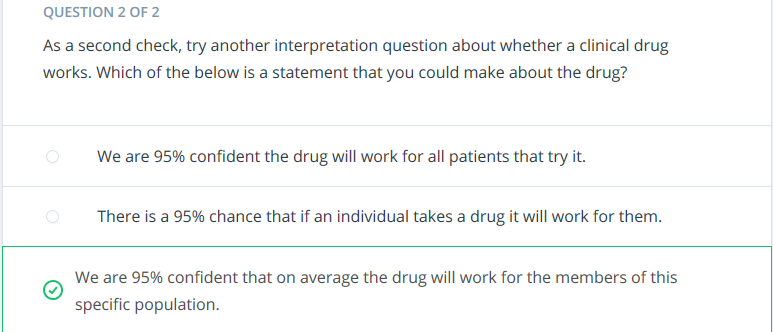




**Confidence Intervals (& Hypothesis Testing) vs. Machine Learning**

* Confidence intervals take an aggregate approach towards the conclusions made based on data, as these tests are aimed at understanding population parameters (which are aggregate population values).
* Alternatively, machine learning techniques take an individual approach towards making conclusions, as they attempt to predict an outcome for each specific data point.





**Recap**

In this lesson, you learned:

1. How to use your knowledge of bootstrapping and sampling distributions to create a confidence interval for any population parameter.
2. You learned how to build confidence intervals for the population mean and difference in means, but really the same process can be done for any parameter you are interested in.
3. You also learned about how to use python built-in functions to build confidence intervals, but that these rely on assumptions like the Central Limit Theorem.
4. You learned about the difference between **statistical significance** and **practical significance**.
5. Finally, you learned about other language associated with confidence intervals like **margin of error** and **confidence interval width**, and how to correctly interpret your confidence intervals. Remember, confidence intervals are about **parameters** in a population, and not about individual observations.

