CONFIDENCE INTERVALS POPULATION MEANS

Unit 4A - Statistical Inference Part 1

UF UNIVERSITY of FLORIDA

{Confidence Intervals for Population Means}

Now we will discuss a few loose ends.

Before moving into our final discussion of confidence intervals for one population mean, let's review a few important results we have for confidence intervals so far.

Consider holding all values constant except the value of interest

We have seen that as n increases, the length of our confidence interval decreases, the confidence interval will be more narrow.

And that as the confidence level increases, the length of our confidence interval increases, the confidence interval will be wider.

Finding the Sample Size needed

$$z^* \cdot \frac{\sigma}{\sqrt{n}}$$

m

$$n = \left(\frac{z^* \sigma}{m}\right)^2$$

- Always choose the next highest integer

UF UNIVERSITY of FLORIDA

We begin with determining the sample size required to estimate the population mean with a specified margin of error.

To find the sample size required to estimate an unknown population mean with a specified margin of error we can algebraically solve for n in our margin of error equation to obtain the equation on the right.

This will give the minimum sample size required. The result is not usually an integer so for decimal results we always take the next highest integer instead of using standard rounding practices.

3

Example

$$n = \left(\frac{z^* \sigma}{m}\right)^2 = \left(\frac{2.576(15)}{2}\right)^2 = 373.26$$

UF UNIVERSITY of FLORIDA

Here is an example.

IQ scores are known to vary normally with a standard deviation of 15.

How many students should be sampled if we want to estimate the population mean IQ at 99% confidence with a margin of error equal to 2?

When we complete the calculation, we get 373.26 students. We cannot get 0.26 of a student and 373 students wouldn't quite be enough so at a minimum we need to take a sample of 374 students.

In practice, sigma is estimated based on the standard deviation obtained in prior studies or a conservative best guess.

4

	Small sample size	Large sample size
Variable varies normally	✓	✓
Variable doesn't vary normally	✗	✓

We also must consider when can we use this approach. In other words, when will the sampling distribution of x-bar be approximately normal? We know from the central limit theorem that we can use the normal approximation whenever the sample size is large enough. For our purposes, we say that n must be larger than 30. When the sample size is large, it does not matter whether the original population is normal or not.

However, for small samples, the sampling distribution of x-bar will ONLY be normal, if the original population was normal. Thus we can only apply this result for small samples when the variable under study is normally distributed in the population. Thus for small samples from non-normally distributed populations, we cannot apply this method.

The conditions under which methods should be applied are important to consider. Many methods rely on approximations, such as those given by the central limit theorem. It is the statistician's goal is to provide accurate methods, even when approximations are utilized. For confidence intervals, this means the statistician would like the actual confidence level to be very close to that claimed by the method (95%, 90%, etc.).

If you use a method when it is not appropriate, the effect on the true confidence level may be difficult to determine and you may be claiming to use a 95% confidence interval when in fact you should only be 85% confident! In other words, the chances that the process will work are smaller than you claim!

5

t-Intervals for a Population Mean

- Assumes the population is normally distributed (or the sample size is large)
- Used when σ is unknown
- t-distribution has degrees of freedom! In this case, they are n-1
- http://www.stat.tamu.edu/~west/applets/tdemo.html

$$\bar{x} \pm t^* * \frac{s}{\sqrt{n}}$$

UF UNIVERSITY of FLORIDA

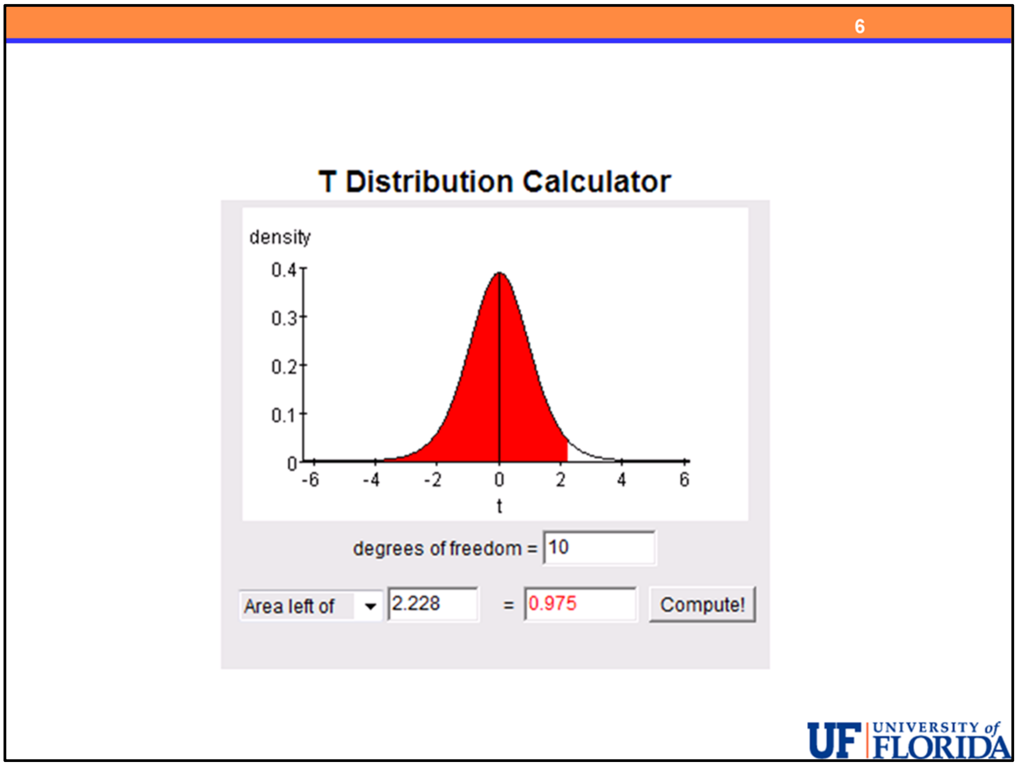
{http://www.stat.tamu.edu/~west/applets/tdemo.html}

We won’t do much with this until we get to case CQ and discuss paired-difference problems, however, we will mention here that there are issues with the central limit theorem when we must substitute the sample standard deviation for the population standard deviation.

Even when the population is normal, this substitution results in the appropriate sampling distribution being a t-distribution with n-1 degrees of freedom rather than a normal distribution.

Although there are online calculators and printed tables for the t-distribution, we will rely on our software to obtain confidence intervals for us in practice.

The formula for the confidence interval is similar except that we have replaced the population standard deviation, sigma, by the sample standard deviation, s, and we have replace the z* value with a corresponding t* value from the appropriate t-distribution with n-1 degrees of freedom.



The t-distribution, although similar to the normal distribution, has heavier tails, there is a greater chance of having a value farther from the center.

As a comparison, for the 95% confidence interval using a normal distribution, we found z* = 1.96.

For a sample size of 11, which would give n – 1 = 10 degrees of freedom, we would obtain t* = 2.228 for the same level of confidence.

As the degrees of freedom increase, the values from the t-distribution and the normal distribution get closer together and the difference between the two methods becomes increasingly negligible.

That concludes our discussion of confidence intervals for a single population mean.

You will be asked to calculate these intervals by hand when the normal distribution can be applied to find the confidence multiplier z*. You will be provided with information about the population standard deviation, the sample size, and the sample mean.

In the case where sigma, the population standard deviation is unknown, we will use software to calculate the confidence intervals based upon the raw data, using the t-distribution.

You should also be able to interpret confidence intervals correctly in context, calculate the required sample size for a specified margin of error, and understand the how the width or margin of error of our interval changes as the confidence level or sample size change.