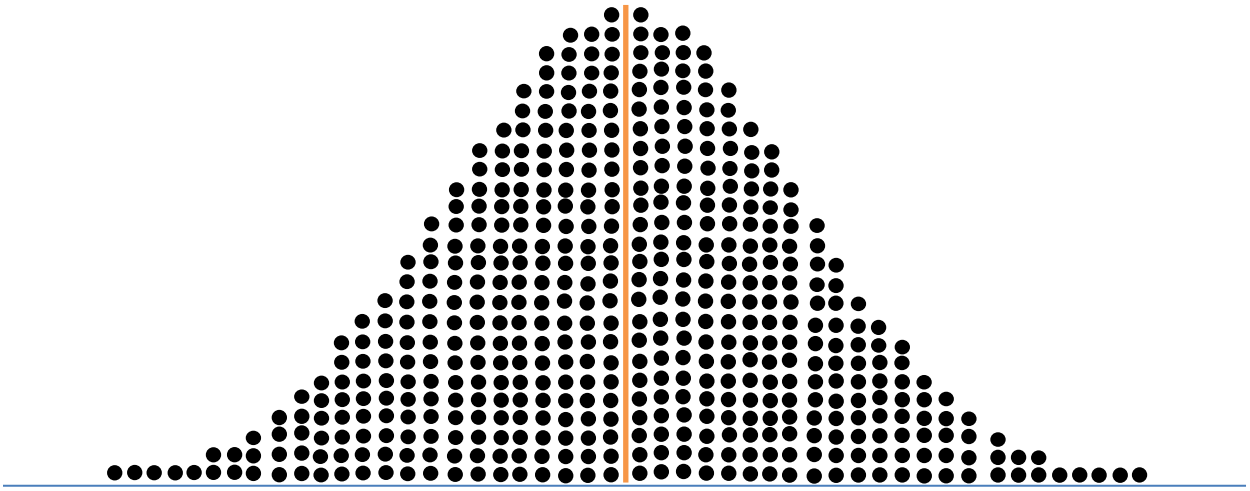
David Youngberg
BSAD 210—Montgomery College

LECTURE 09: CONFIDENCE INTERVALS I

I. The Meaning of Confidence
   a. When we take a sample, we often summarize it with the sample's mean. This is an example of a *point estimate*—a single value that best describes the population. The problem is, our sample may be unusual.
   b. Suppose the population mean is the vertical orange bar. We don't know it, but we know that, by the central limit theorem, the sample means will form a normal distribution around it. Each black dot represents a sample mean we could get when we take our sample.
   c. Most sample means are right by the orange line. Some are a bit far away. A few are quite far away.



   d. We are trying to figure out a good way to use the information of the point estimate while recognizing our point estimate is likely not right on target. We do this with a confidence interval.
      i. A *confidence interval for the mean* is the range where the true population mean lies.
   e. Imagine a horizontal line, like this:

      i. This is your confidence interval. Imagine centering that purple line on various black dots above. When that line crosses the orange line, then the population mean is in the interval. When it doesn't, then it's not in the interval.

ii. Every purple line has an associated *confidence level*, or the probability that the interval estimate will include the population mean (μ).
iii. For example, if this purple line was for 95% confidence, that means when centered on 95% of the black dots, the purple line would include the population average. Since we "pick" one of the dots at random, we can say that such an interval has a 95% chance of including the population mean.
iv. If we wanted there to be a 99% chance of getting the population mean? That would naturally require a larger range.

II. Z-scores

a. How much larger? That's where z-scores come in; z-scores are basically standard deviations. That means at a z-score of 2, 95% of observations are within two standard deviations. Every confidence rating has a z-score associated with it.
   i. Technically, 95% confidence is a z-score of 1.96. We'll use these numbers when we get to calculating confidence intervals.
b. There are three standard confidence levels: 95%, 99%, and 99.9%. Which should you use?
   i. All are useful. 95% confidence narrows the choices of where the parameter lies but 99.9% confidence will have a large range and thus you'll be more likely to include the parameter. 99% is a mix between then.
c. The Greek letter α (alpha) is the *significance level*; it's equal to 1 – confidence level. It's everything the confidence level isn't. If 95% is the chance that your interval includes the population mean, 5% (or 0.05) is the chance it doesn't.
   i. For some reason, confidence intervals are all percents and significance levels are all decimals.
   ii. We tend to refer to α as the tail probability, because it's the least likely part—it happens in the "tail" or "tails" of the distribution.
   iii. Because there are two sides of the normal distribution, α is sometimes divided by two to indicate the area for alpha is split in half: $\alpha/2$.
   iv. Every critical value has a corresponding z-score, $z_{\alpha/2}$. Below is a table of regularly used z-scores.

Confidence	α	$z_{\alpha/2}$
95%	0.05	1.960
99%	0.01	2.576
99.9%	0.001	3.291

v. As a life skill, it's useful to memorize this list even if you only memorize it out to two decimal places rather than three.

III. Calculating the Margin of Error with Known σ

a. The *margin of error* determines the width of a confidence interval; it is the distance between an upper or lower bound and the mean.
   i. Because the normal distribution is symmetric, the distance between the upper bound and the sample mean equals the distance between the lower bound and the sample mean.
   ii. Twice the margin of error is the range of the confidence interval.

b. When the population standard deviation (σ) is known, the equation for a confidence interval is this:

$$CI_{\bar{x}} = \bar{x} \mp z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

   i. Where $CI_{x\text{-bar}}$ is the confidence interval for the sample mean;
   ii. *x-bar* is the sample mean;
   iii. $z_{\alpha/2}$ is the critical value for α significance level;
   iv. σ is the population's standard deviation; and
   v. *n* is the sample size.
   vi. Note the minus/plus sign means you subtract (to the get the lower bound) and add (to get the upper bound).
   vii. The margin of error is everything after the minus/plus sign.

c. You typically know σ concerning well-established data. This includes blood pressure, stock market prices, or any other data that's calculated regularly and for a long period of time.

d. But knowing σ means you know μ (recall you need μ to calculate σ) so a confidence interval isn't as useful here.
   i. Sometimes μ changes and we have good reason to think σ didn't change—like shifting the entire normal distribution right or left—so this there's still a use for what we just covered.

IV. Confidence in Excel

a. Excel's confidence function calculates the margin of error; you subtract from and add to the mean yourself.

b. The good news is that Excel has all z-scores for *any* alpha built in. So not only do you just need to supply the alpha, you are not limited to the 0.05, 0.01, and 0.001 alphas.

c. The command is "=CONFIDENCE.NORM"
   i. *Alpha* is the significance level, expressed as a decimal.
   ii. *Standard_dev* is the standard deviation of the population

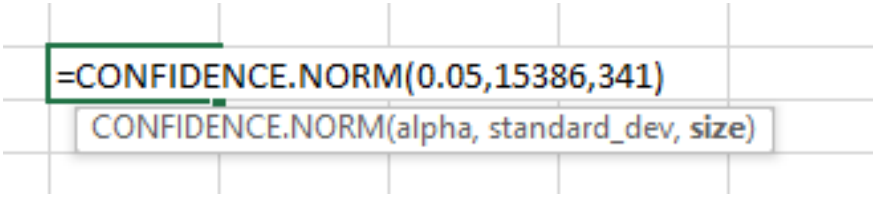
iii. *Size* is the sample size.

V. Example

a. Theo works at Friendly Friends, a toy company. Every few years the company makes slight alterations to its teddy bear (it's their best selling product). While most changes increase sales and a few decrease sales, the standard deviation of sales is always more or less the same: 15,386.

b. The company just made a change to the bear and Theo needs to figure out how many bears to make for the holiday season. He first runs several focus groups. Based on the input of 341 people, he estimates they will sell an average of 413,512 bears. At 95% confidence, what's the confidence interval?

c. Type "=CONFIDENCE.NORM(0.05,15386,341)" into Excel and press ENTER. You should get about 1,633.



i. Therefore, they should make between 411,879 and 415,145 bears. If they want to make sure they have enough bears in stock, the upper limit is where they want to be.

d. Now change the alpha value to a cell reference (I used A2, as shown) and put 0.05 there. Then add 0.01 in the cell below that and 0.001 in the cell below that.

	A	B
1	Alpha	Margin of Error
2	0.05	1633.039
3	0.01	
4	0.001	

	A	B
1	Alpha	Margin of Error
2	0.05	1633.039
3	0.01	2146.177
4	0.001	2741.662

e. Note that as the significance level falls (and thus the confidence level increases), the greater the margin of error.

i. This highlights the conflict when choosing a significance level. The lower the level, the better chance you will capture the true mean. But the bigger the margin of error, the vaguer the confidence interval.