David Youngberg
BSAD 210—Montgomery College

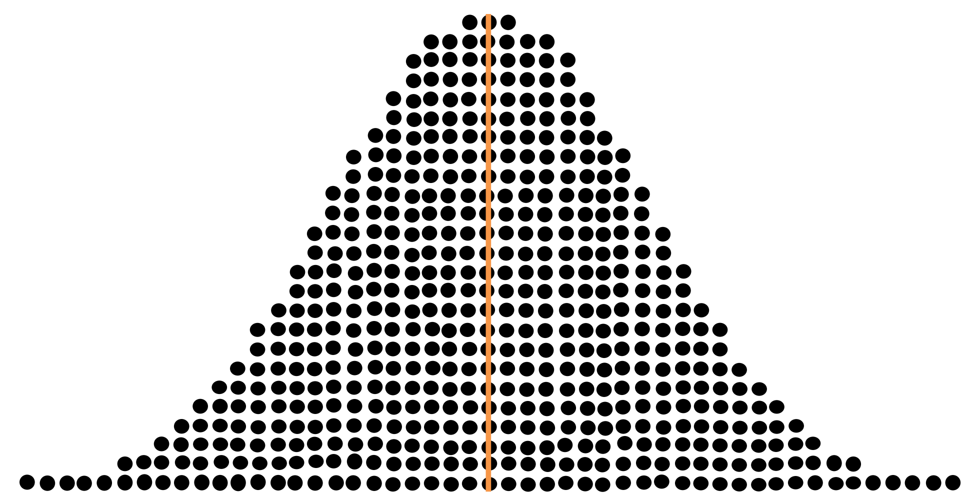
LECTURE 13: CONFIDENCE INTERVALS I

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-score*s*

a. Z-scores are critical to confidence intervals. Indeed, we call them *critical values* because that is the threshold value. For example, what should have a larger range: 95% confidence or 99% confidence?

i. Both are useful. 95% confidence narrows the choices of where the true parameter lies but 99% confidence will have a larger range and thus you'll be more likely to include where the true value is.

ii. The question becomes, how do you mathematically capture that change from 95% to 99%?

b. 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.

c. The Greek letter α (alpha) is the *significance level*; it's equal to 1 – confidence level. For example, at 95% confidence, $\alpha = 0.05$.

i. 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: $\alpha/2$.

ii. 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

iii. 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.

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 a steel mill and is in charge of making sure there's enough scrap metal in stock to keep the mill operating, but not so much that it takes up too space and ties up too much money.

b. Theo's job is tricky because scrap metal prices vary widely and production varies wide. If prices are high, Theo knows to buy less but *how much* less? He doesn't want to buy so little, the mill runs out of scrap metal for important jobs. If prices are low, Theo knows to buy more but *how much* more? He doesn't want to buy so much, the mill sits on it for months and it rusts as it takes up room.

c. What Theo would like is a range of production. You can't predict how much scrap iron the mill uses weeks in advance, but a range of expectations would be very helpful. This is where a confidence interval comes in.

d. Theo knows the company uses, on average, 7,000 tons a week (based on the previous 20 weeks of production), with a population standard deviation of 800 tons. At 95% confidence, what's the confidence interval?

e. Type "=CONFIDENCE.NORM(0.05,800,20)" into Excel and press ENTER. You should get about 350.6 tons.

    i. Add and subtract this number from 7,000. Theo can predict, at 95% confidence, that the mill will use between 6,649.4 to 7,350.6 tons and he can keep these numbers in the back of his mind as he's purchasing scrap metal.

f. What if Theo fears his 95% range doesn't capture the scrap metal use that week? After all, 95% confidence means there's a 5% chance the true population mean is outside of that range. One way to solve this problem is to increase the confidence level to 99% or 99.9% (alpha = 0.01 and 0.001, respectively) as so:

	A	B
1	**Alpha**	**Margin of Error**
2	0.05	350.6
3	0.01	460.8
4	0.001	588.6

g. 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.