EPSE 482: The Central Limit Theorem.

Oscar L. Olvera Astivia, Ph.D.

The University of British Columbia

oolvera@mail.ubc.ca

Feb 5, 2019

From last class...

- We learned what are probability distributions and, more importantly, the very special case of the normal distribution.
- We learned about the μ and σ parameters of the normal distribution and how it changes the shape of it.
- We learned how to calculate the probabilities under the normal curve using the 68—95—99.7 rule.
- We learned how to standardize scores and transform them from one metric to another.

Inferential statistics

We collect a *sample* of data from a *population* because it usually unfeasible to access all the elements of the population.

Inferential statistics

We collect a *sample* of data from a *population* because it usually unfeasible to access all the elements of the population.

The overall gist of what we do relies on the assumption that our sample statistics are the best estimates we have of the population parameters. Sample statistics are our best "guess" of that the population value is.

Inferential statistics

We collect a *sample* of data from a *population* because it usually unfeasible to access all the elements of the population.

The overall gist of what we do relies on the assumption that our sample statistics are the best estimates we have of the population parameters. Sample statistics are our best "guess" of that the population value is.

Most of the data analysis methods that you will be using (and learning about) rely on the 2 most important mathematical results there are in the world of Statistics: The Law of Large Numbers and the Central Limit Theorem.

The Law of Large Numbers

I find this one to be the easiest one to understand because what it says is relatively intuitive:

In plain English: As your sample size grows larger and larger, the sample mean $\bar{x}$ converges to its population value μ (***).

The Law of Large Numbers

I find this one to be the easiest one to understand because what it says is relatively intuitive:

In plain English: As your sample size grows larger and larger, the sample mean $\bar{x}$ converges to its population value μ (***).

In Math:

Definition

Let X be a real-valued random variable, and let $X_1, X_2, X_3, ...$ be independent and identically distributed (i.i.d.) realizations of X. Define $\bar{X}_n = (1/n)(X_1 + X_2 + X_3 + ... + X_n)$. Then:

$$Pr(\lim_{n \to \infty} \bar{X}_n = \mu) = 1$$

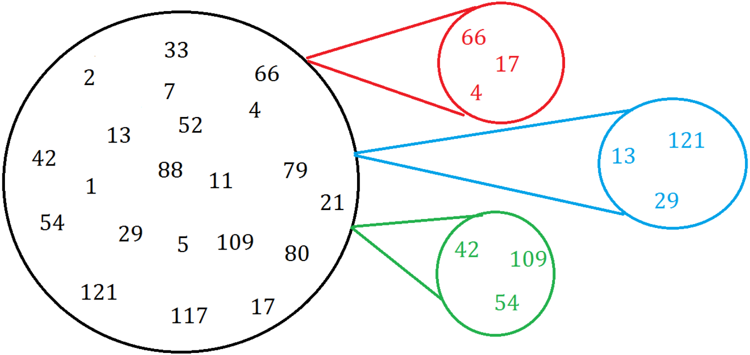
The Law of Large Numbers

The sampling distribution: The basics

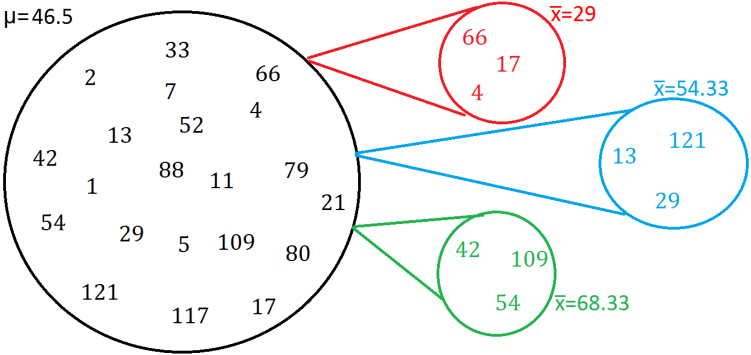
Until now we have been rehashing ideas of random variables, probability distributions (like the normal) and sample statistics (or sample estimates) VS population parameters. We are tying this up all now:

Your **sample statistics** are **random variables** that follow a **probability distribution**.

The sampling distribution: The basics



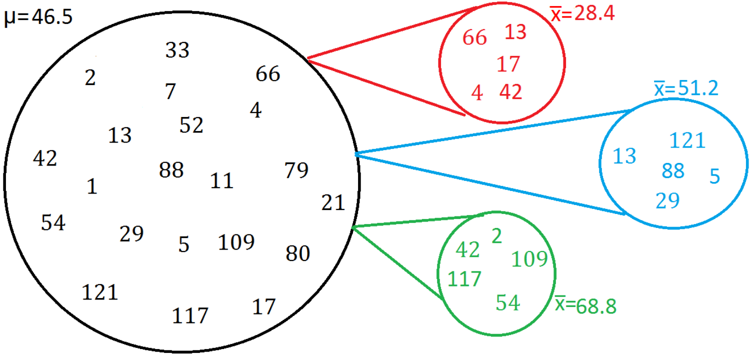
The sampling distribution: The basics



The sampling distribution: The basics

As we can see, every sample average for sample sizes $n = 3$ is different. Some are going to be higher and some are going to be smaller than 46.5. Let's see what happens if I increase my sample sizes just a tad little bit, from 3 to 5.

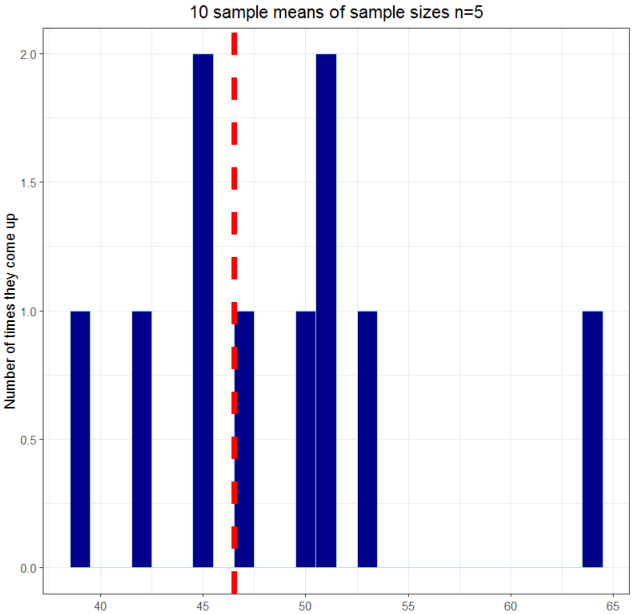
The sampling distribution: The basics



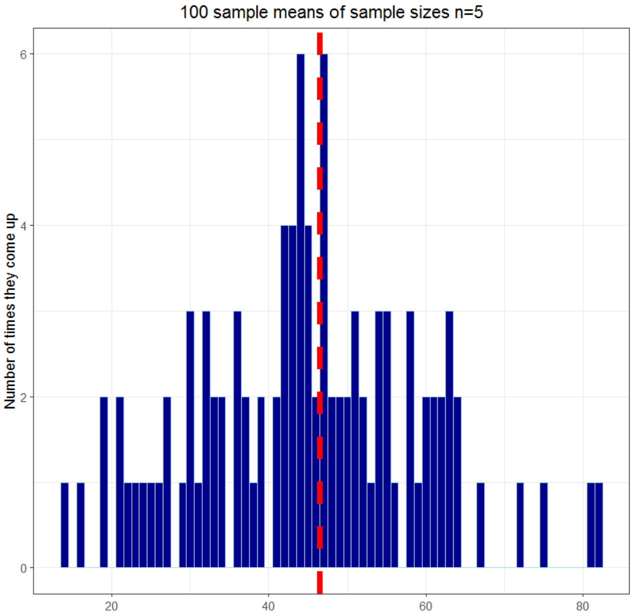
The sampling distribution: The basics

What we are now going to do is what underlies the engine of most of inferential statistics. We are going to keep on taking *repeated samples* of size $n = 5$ from our population. Let's plot them to see if we find any pattern:

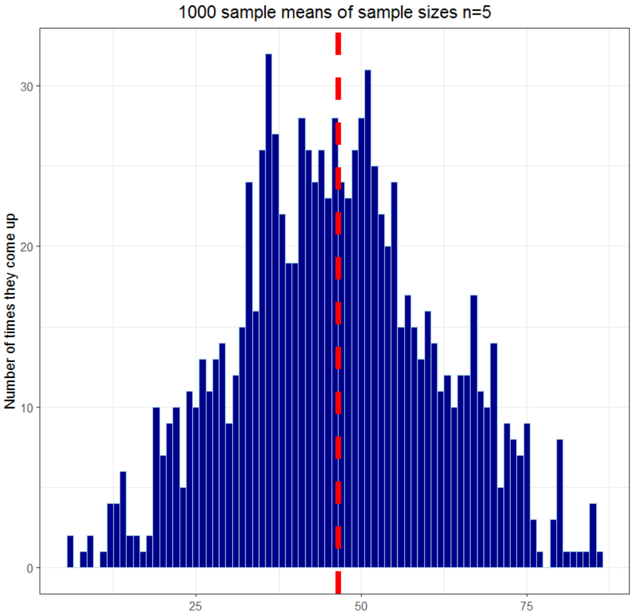
The sampling distribution: The basics



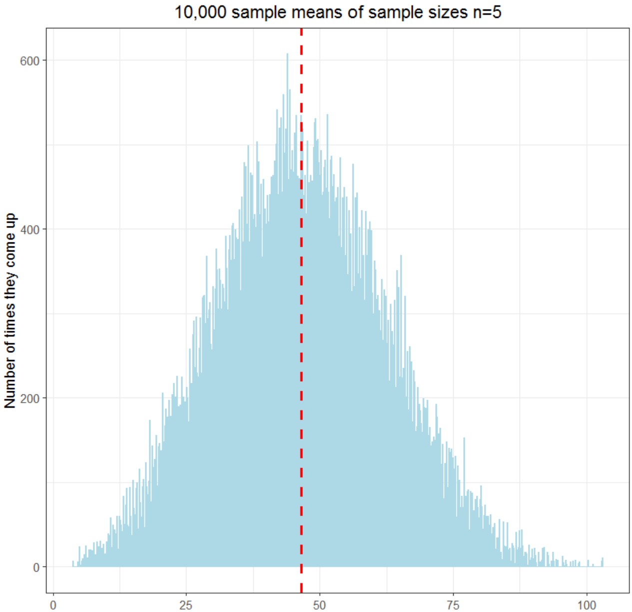
The sampling distribution: The basics



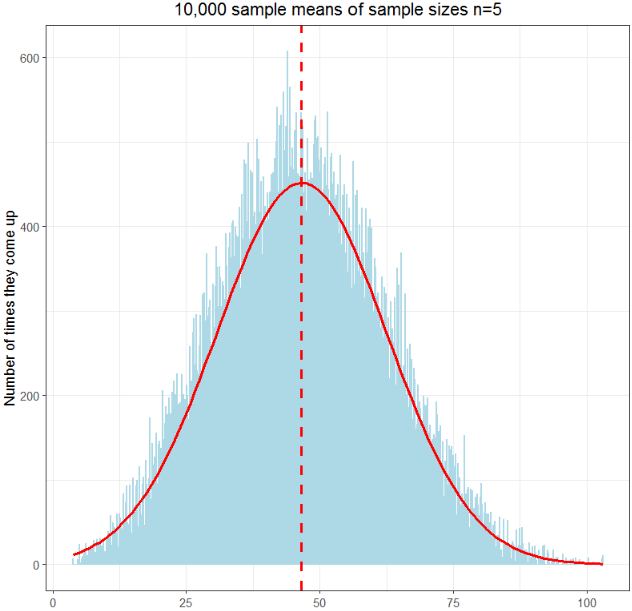
The sampling distribution: The basics



The sampling distribution: The basics



The sampling distribution: The basics



The sampling distribution: The basics

Even though our sample sizes are small ($n = 5$), if we keep on taking more and more samples of size 5 and plotting them, the resulting histogram is going to resemble more closely a normal distribution.

The sampling distribution: The basics

Even though our sample sizes are small ($n = 5$), if we keep on taking more and more samples of size 5 and plotting them, the resulting histogram is going to resemble more closely a normal distribution.

This is the gist of what we call in statistics the **Central Limit Theorem** which is that if you take an infinite number of samples of any size from some population, take the sample averages and plot them, you find that the *sampling distribution* of these averages follows a normal distribution.

The Central Limit Theorem

A **sampling distribution** is the probability distribution that a statistic (i.e. a random variable) follows after a large number of random samples is drawn from the same population.

The Central Limit Theorem

A **sampling distribution** is the probability distribution that a statistic (i.e. a random variable) follows after a large number of random samples is drawn from the same population.

The **Central Limit Theorem** is the mathematical result that guarantees that the sampling distribution of the mean is normal, IRRESPECTIVE of the shape of the parent population.

The Central Limit Theorem

Central Limit Theorem (CLT)

Let $\{X_1, \ldots, X_n\}$ denote a random sample of n independent observations from a common distribution with finite mean μ and finite variance σ^2. Recall the sample mean is given by:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Then, for n large, X is approximately distributed as $N(\mu, \sigma^2/n)$.

The Central Limit Theorem

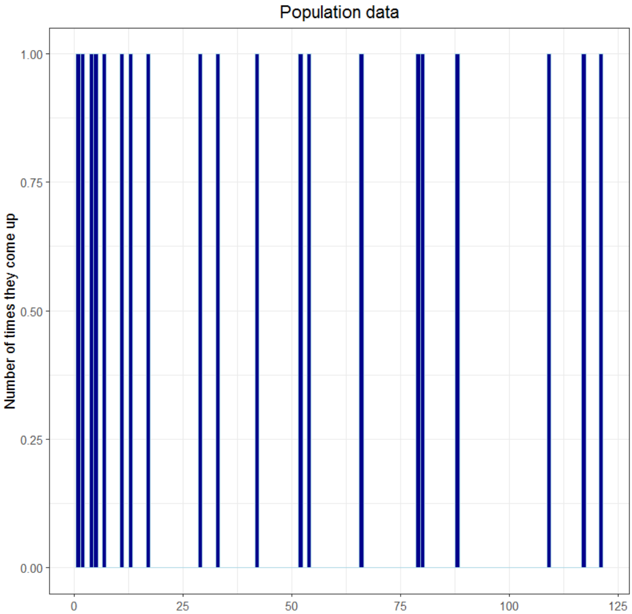
The power behind the Central Limit Theorem is due to its generality. It doesn't matter what the shape of the population distribution is, it doesn't matter if it's discrete or continuous, etc. For a sufficiently large n any sample $\bar{x} \sim \mathcal{N}(\mu, \sigma^2/n)$

The Central Limit Theorem

The power behind the Central Limit Theorem is due to its generality. It doesn't matter what the shape of the population distribution is, it doesn't matter if it's discrete or continuous, etc. For a sufficiently large n any sample $\bar{x} \sim \mathcal{N}(\mu, \sigma^2/n)$

Look at the "population" from our example! How does it look like?

The Central Limit Theorem



That's right! My "population" shape is not even continuous! It really is just a collection of numbers with a 1/20 probability of being selected. So it does not really matter what shape or type the probability distribution of your random variables follow. If you:

1. Take a sample from ANY population.
2. Calculate the sample average $\bar{x}$ from this sample.
3. Save that sample $\bar{x}$ in a list somewhere.
4. Go back to Step 1 and repeat many, many, MANY times.

When you plot all your sample $\bar{x}$, the curve that results is a normal distribution.

The Central Limit Theorem: μ and σ

So we know now that a normal distribution has 2 parameters that control the location and the spread: μ and σ.

The sampling distribution of the sample mean $\bar{x}$ is also normal, so, what is its μ and σ?

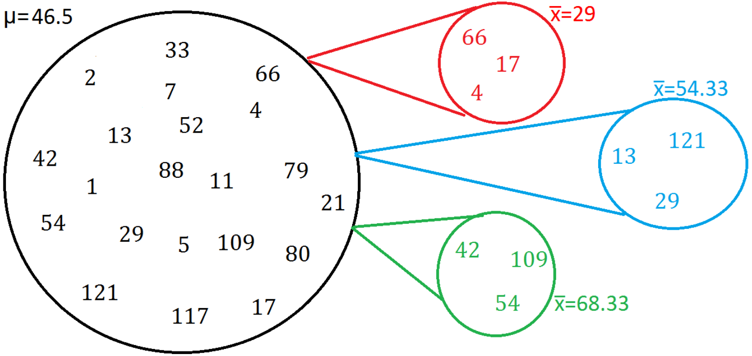
The Central Limit Theorem: μ and σ

So we know now that a normal distribution has 2 parameters that control the location and the spread: μ and σ.

The sampling distribution of the sample mean $\bar{x}$ is also normal, so, what is its μ and σ?

The first one is a little intuitive (I hope). This "average of averages", for a large enough n, is actually the original population's μ.

The Central Limit Theorem: μ



The Central Limit Theorem: μ

We know the true population μ is 46.5

Now, if we do:

$$\frac{29 + 54.33 + 68.33}{3} = 50.55$$

which is closer to 46.5 than any individual sample!

Remember! True population $\mu = 46.5$

Number of samples $n = 3$	"Average of averages"
1	50.55
10	48.62
100	43.77
1000	46.48
10,000	46.55

The Central Limit Theorem: μ

I hope this helps convey the intuition of why the more samples we gather, the closer we get to our population parameter estimates.

The Central Limit Theorem: μ

I hope this helps convey the intuition of why the more samples we gather, the closer we get to our population parameter estimates.

In a way, the fact that we are constantly averaging over averages implies that we are cancelling out the "noise" so that we can extract the "signal".

The Central Limit Theorem: μ

I hope this helps convey the intuition of why the more samples we gather, the closer we get to our population parameter estimates.

In a way, the fact that we are constantly averaging over averages implies that we are cancelling out the "noise" so that we can extract the "signal".

As you can expect, in general, large samples are better than small samples (Law of Large Numbers) and many samples are better than a few samples (Central Limit Theorem).

The Central Limit Theorem: σ

Recall that:

$$\frac{29 + 54.33 + 68.33}{3} = 50.5533$$

There is nothing stopping us from calculating the standard deviation of those 3 values like:

$$s^2 = \frac{(29 - 50.55)^2 + (54.33 - 50.55)^2 + (68.33 - 50.55)^2}{3 - 1}$$

$$s^2 = \frac{794.81}{2}$$

$$s = \sqrt{397.40} = 19.93$$

The Central Limit Theorem: σ

Remember! The sample mean $\bar{x}$ *is* a random variable. And, as such, its value can change with various degrees of probability (i.e. area under the normal curve). We need some way to calculate the uncertainty around that sample mean $\bar{x}$. The standard deviation (or the variance) can do the trick for us.

The Central Limit Theorem: σ

Number of samples $n = 3$	"Average of averages"	Standard deviation
1	50.55	19.93
10	48.62	15.47
100	43.77	14.10
1000	46.48	15.41
10,000	46.55	15.85

The Central Limit Theorem: σ

Although our "average of averages" gets closer and closer to the true population mean of 46.5, there is still quite a bit of variability because the sample size is so small, $n = 3$

What would happen if instead gathering many samples of size $n = 3$ we increased our samples to say $n = 10$

The Central Limit Theorem: σ

Number of samples $n = 10$	"Average of averages"	Standard deviation
1	47.6	46.65
10	47.12	6.35
100	46.04	9.51
1000	46.46	8.95
10,000	46.54	9.07

The Central Limit Theorem: σ

Notice how increasing the sample size decreases the standard deviation $n = 3$ vs $n = 10$, but increasing the number of samples taken from the population doesn't immediately mean that the standard deviation gets reduced.

For the case of $n = 3$, it hovers around 14 to 16, and for $n = 10$ it's around 6 to 9. Why?

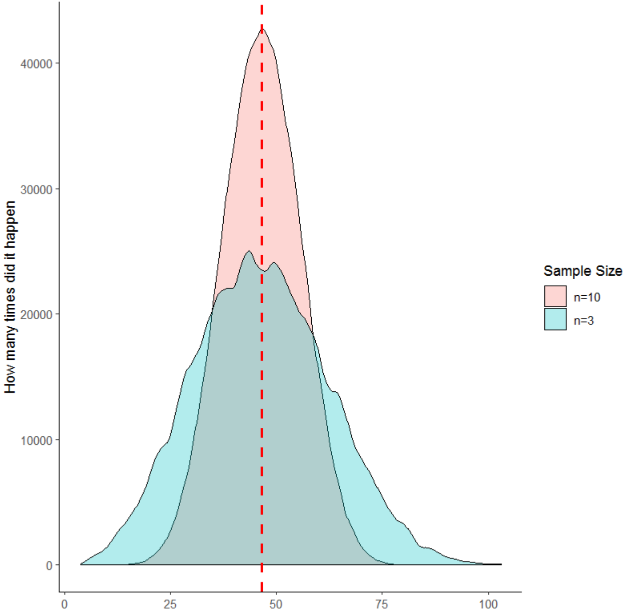
The Central Limit Theorem: σ

Notice how increasing the sample size decreases the standard deviation $n = 3$ vs $n = 10$, but increasing the number of samples taken from the population doesn't immediately mean that the standard deviation gets reduced.

For the case of $n = 3$, it hovers around 14 to 16, and for $n = 10$ it's around 6 to 9. Why?

It mostly has to do with the fact that for every fixed n there is a normal probability distribution associated with it.

The Central Limit Theorem: σ



The Central Limit Theorem: σ

As the sample size becomes larger and larger, there is more and more information about where the population value μ is. That is why the standard deviation of the sampling distribution becomes smaller and the normal distribution becomes narrower and narrower. If we could somehow capture this hypothetical infinite population, then the standard deviation of the sampling distribution would be 0.

The Central Limit Theorem: σ

As the sample size becomes larger and larger, there is more and more information about where the population value μ is. That is why the standard deviation of the sampling distribution becomes smaller and the normal distribution becomes narrower and narrower. If we could somehow capture this hypothetical infinite population, then the standard deviation of the sampling distribution would be 0.

Because we need the parameter σ to reflect that this increase in information should reflect a decrease in uncertainty, the standard deviation of the sampling distribution is defined as either $\boldsymbol{\sigma^2/n}$ or $\boldsymbol{\sigma/\sqrt{n}}$

Central Limit Theorem: Joining it all together

Remember: we rarely have access to the population parameters (μ, σ). We almost always exclusively operate with the sample statistics $(\bar{x}, s)$. Because of this, our "best guess" for population μ is sample $\bar{x}$ and for population σ is s.

Central Limit Theorem: Joining it all together

Remember: we rarely have access to the population parameters (μ, σ). We almost always exclusively operate with the sample statistics $(\bar{x}, s)$. Because of this, our "best guess" for population μ is sample $\bar{x}$ and for population σ is s.

But, REMEMBER! the uncertainty around $\bar{x}$ needs to reduce as n grows larger and larger. Which is why, to calculate its variability we are going to use one of the most important statistics you will ever encounter: the **STANDARD ERROR OF THE MEAN**:

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

Central Limit Theorem: Joining it all together

The **standard error of the mean** is our "best guess" at what the standard deviation of the sampling distribution is. If you think about it, the fact that you can have good "one shot" estimate of something that, in theory, requires an infinite number of repetitions is quite remarkable!

Central Limit Theorem: Joining it all together

Let's pretend for a moment that our "population" is the numbers from 1 to 1000. Yes, so $S = \{1, 2, 3, 4, 5, ..., 98, 99, 100, 101, 102, ..., 998, 999, 1000\}$

For this population, $\mu = 500.5$ and $\sigma = 288.81$. Let's start taking samples of different sizes (notice NO repeated sampling) and compare $\sigma/\sqrt{n}$ to $s/\sqrt{n}$

Central Limit Theorem: Joining it all together

Sample size	$\bar{x}$	σ^2/n	$SE(\bar{x})$
10	576.20	91.33	92.81
50	527.12	40.84	42.16
100	509.37	28.88	29.92
500	496.13	12.91	12.99
1000	500.5	9.13	9.13

Central Limit Theorem: Joining it all together

Two important questions:

(1) Why is the standard error of the mean not 0?

Two important questions:

(1) Why is the standard error of the mean not 0?

Well, remember that the assumption is that it would only become 0 as $n \to \infty$. The formula was only derived under the more general case that you have an *infinite* population. And 1 to 1000 is large... but NOT infinite! Still, you can use the result and see that $SE(\bar{x})$ approximates $\sigma/\sqrt{n}$ really well. And why is that? (Hint: Law of Large Numbers).

(2) What is the difference between the standard error of the mean and the standard deviation?

Central Limit Theorem: Joining it all together

(2) What is the difference between the standard error of the mean and the standard deviation?

This one is perhaps one of the things that stumps people the most. The concepts are related but they ARE NOT the same.

The **standard deviation** σ is a population parameter that pertains to the variability of INDIVIDUAL data points around the population parameter μ

The **standard error of the mean** $SE(\bar{x})$ is a sample statistic that approximates a population parameter (i.e. the standard deviation of the sampling distribution) and pertains to the variability of small sample averages $\bar{x}$ around the population parameter μ.