The Central Limit Theorem (Section 9.2)

The Central Limit Theorem. Provided the conditions below are met (large sample or population distribution is normal) then *the distribution of sample means will be approximately* ***normal***.

Conditions that must be met **for Central Limit Theorem to apply** (page 416):

1. **Random and Independent.** The sample must be random and the observations in the sample must be independent from one another.
2. **Large Sample or Normal Pop.**

   Either the sample size is "large" ($n \geq 25$)

   OR if you have a "small" sample, then you have to check or assume the underlying population is Normal $\quad \hookrightarrow n < 25$

Features of the Sampling Distribution of the Sample Means:

The mean of all the sample means is the same as the population mean.

$\mu_{\bar{x}} = \mu$ (The mean of the means is the mean…ay yai yai! But easy in practice!)

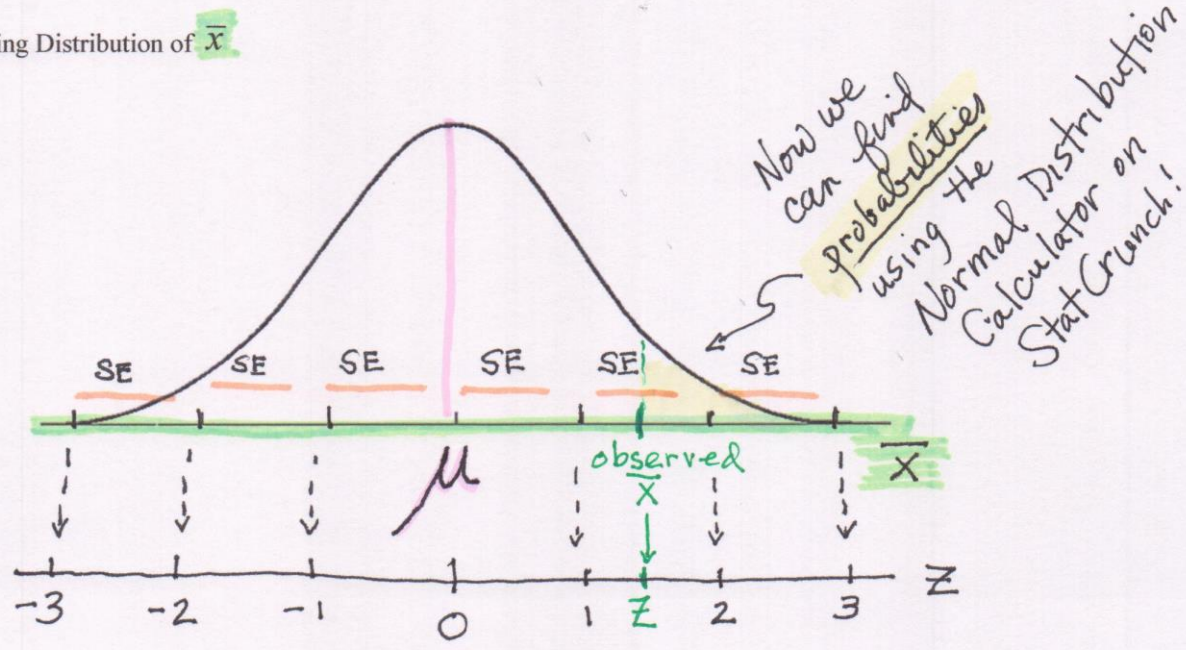
The standard deviation of the all the sample means is called the "Standard Error"

$$\sigma_{\bar{x}} = SE = \sigma/\sqrt{n}$$

The z-score for an observed Sample Mean, $\bar{x}$, is found the usual way: $z = \dfrac{observed - center}{SD}$

In this context, this is the formula: $\boxed{z = \dfrac{\bar{x} - \mu}{SE}}$ ~ or $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Sampling Distribution of $\bar{x}$



Example: Air Pollution and Children's Health. One of the measurements used to determine the health of a person's lungs is the amount of air a person can exhale under force in one second. This is called the "forced expiratory volume in one second" and is abbreviated FEV_1. Previous studies have established that the mean FEV_1 for 10-year-old boys is 2.10 liters and that the population standard deviation is 0.3 liters. A random sample of 100 10-year-old boys who live in a community with high levels of ozone pollution is found to have a mean FEV_1 of 1.95 liters.

What is the probability that the sample of boys in the polluted of areas would have a mean of ~~2.02~~ 1.95 liters (or less) if the true population average for boys in polluted areas was actually 2.10 liters?

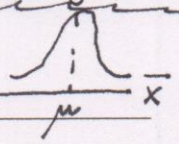
What do we know? (Parking Lot)

$\mu = 2.1$ L
$\sigma = .3$ L
$n = 100$
$\bar{x} = 1.95$ L

$SE = \frac{\sigma}{\sqrt{n}} = \frac{.3}{\sqrt{100}} = .03$

What do we want to find out?

$P(\bar{x} \le 1.95)$
= how likely is it the sample mean would be this small if the boys in polluted area were the same, on average, as all 10-year old boys?



Strategy: To find the probability we'll use the Central Limit Theorem

First, we have to check that the conditions are met by this problem.

Check conditions for CLT:

1. Random sample? – yes, stated
   Independent observations? Assume 1 boy's FEV_1 score doesn't affect another's – no connection
2. Large sample OR Small with Normal Pop?
   $n = 100 \ge 25$
   Yes, large sample

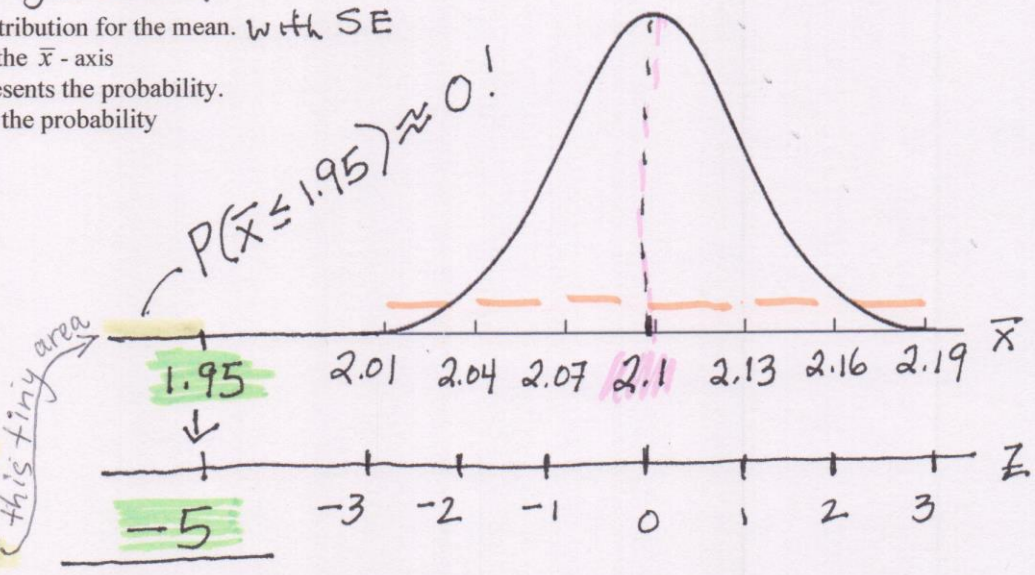
Graph the Sampling Distribution for the mean. with SE
Sketch the z-axis under the $\bar{x}$ - axis
Shade the area that represents the probability.
Find the the z-score and the probability

$$z = \frac{\bar{x} - \mu}{SE}$$

$$= \frac{1.95 - 2.1}{.03}$$

$$= -5$$

$P(\bar{x} \le 1.95)$
$= P(z \le -5)$
$= .00000029$



If this were a hypothesis test, what would we call this probability?

The P-value

Types of significance:

"Evidence-based practice requires clinicians to stay current with the scientific literature. Unfortunately, rehabilitation professionals are often faced with research literature that is difficult to interpret clinically. Clinical research data is often analyzed with traditional statistical probability (p-values), which may not give rehabilitation professionals enough information to make clinical decisions. **Statistically significant** differences or outcomes simply address whether to accept or reject a null or directional hypothesis, without providing information on the magnitude or direction of the difference (treatment effect). To improve the interpretation of **clinical significance** in the rehabilitation literature, researchers commonly include more clinically-relevant information such as confidence intervals and effect sizes. … Understanding these aspects of research will help practitioners better utilize the evidence to improve their clinical decision-making skills."
Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4197528/

- **Statistical significance:** This is based on an **inference** made by using a hypothesis test.

  Did the kids in the high pollution areas have statistically significant reduction in lung function? Explain.

  Yes! There is essentially an almost ZERO chance that the boys in the polluted area would, on average, have this reduction in FEV_1 in the sample, if the population of ALL boys in this area were, on average, the same as boys in all areas.

- **Clinical significance:** This is determined by experts in health and mental health fields. A *statistically significant* effect may or may or may not be *clinically significant*!

  In the context of the air pollution problem above, do we know how much of a reduction in FEV_1 is ***clinically significant***? Explain.

  No, we do not know whether this reduction is clinically significant because we aren't experts in this field. Pediatricians and pulmonologists would need to analyze this further.

- **Practical significance:** This is determined based on the context of the research and what matters in that situation. Again, a *statistically significant* finding may or may or may not be *practically significant*!

  Climate data shows there has been an increase of average ocean temperatures of statistically significant
  1 degree Fahrenheit over the last century. Is this small change in temperature of any ***practical significance***?

  Small change - does it matter?
  YES - in this context it does!

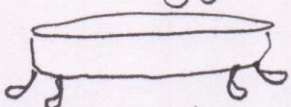
  1°F of heat applied over vast area (volume, really) of the world's oceans = HUGE amount of energy!!

  Energy in …

  Cup of Boiling H_2O

  Energy in …

  Bathtub of boiling H_2O