M358K : October 9th, 2020.

Sample Mean (Normal case).

Say, we are modelling a particular phenomenon using the normal distribution.

That means that $X_1, X_2, \dots, X_n$ are all independent and identically distributed normal.

We write $X \sim Normal(mean = \mu_X, sd = \sigma_X)$

unknown (μ_X) known (σ_X)

w/ X playing the role of "representative" of the X_i, $i = 1..n$.

Consider the sample mean $\bar{X}$:

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

$$\Rightarrow \quad \bar{X} \sim Normal\left(mean = \mu_X, \ sd = \frac{\sigma_X}{\sqrt{n}}\right)$$

or

$$var = \frac{\sigma_X^2}{n}$$

$$\Rightarrow \quad \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \sim N(0,1)$$

University of Texas at Austin

Problem Set # 5

Sample mean: The normal sample.

Problem 5.1. The scores of individual students on the Advanced Dark Arts Exam are modeled as normally distributed with a mean of 19.6 and a standard deviation of 5.0. At Voldemort High, 64 seniors take the test. Assume the individual scores at this school are modeled using the same distribution as national scores. What is the sampling distribution of the sample average score for this random sample of 64 students?

State the **name** and the **parameter value**(s) of this distribution.

$$\bar{X} \sim Normal(mean = \underline{19.6}, \ std\,dev = \underline{0.625})$$

$$= \frac{5}{\sqrt{64}}$$

Problem 5.2. The *"Aristocratic Hog"* chocolate bars are all labeled to weigh 4.0 ounces. The distribution of the actual weights of these chocolate bars is modeled as normal with a mean of 4.0 ounces and a standard deviation of 0.1 ounces. Bernard, the quality control manager and principal taster, initially plans to take (and weigh) a simple random sample of size n from the production line. Then he reconsiders and decides that a sample twice as large is needed. By what factor does the standard deviation of the sampling distribution of the sample average change?

$$SD[\bar{X}_n] = \frac{0.1}{\sqrt{n}}$$

$$SD[\bar{X}_{2n}] = \frac{0.1}{\sqrt{2n}}$$

$$\frac{1}{\sqrt{2}}$$

Problem 5.3. The individual students' scores in the ACT exam are modeled using the normal distribution with an unknown mean (say, it varies from year to year) and with the **known** standard deviation of 6. $\sigma_X = 6$

You take a SRS of students who took the ACT this year. The intention is to use their sample average to estimate (infer) the population mean.

You want the standard deviation of your statistic $\bar{X}_n$ to be at most 0.10. What is the least number of students you need to sample?

$$SD[\bar{X}_n] \leq 0.10$$

$$\frac{6}{\sqrt{n}} \leq 0.10$$

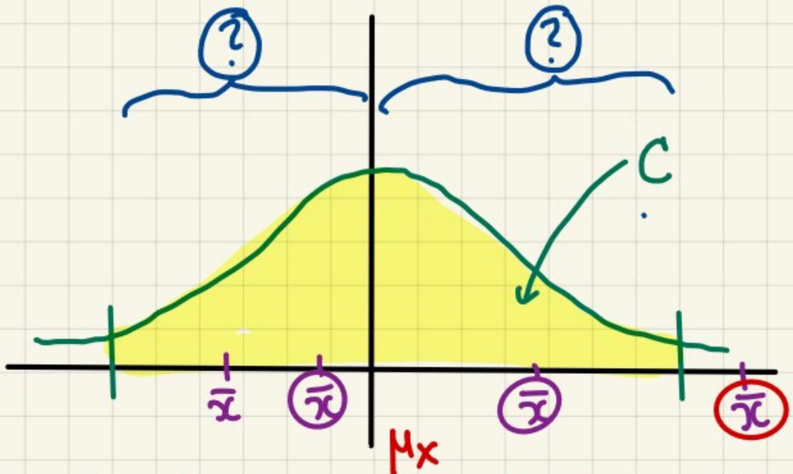
$$60 = \frac{6}{0.1} \leq \sqrt{n}$$

$$\boxed{3600 \leq n}$$

Confidence Intervals

We know that $\overline{X}$ is a "good" point estimator for the population mean μ_X.

Q: How confident are we about the value we get? What does this "confidence" even mean?



Let $C \in (0,1)$ be a ("large") probability.

Say $C = 0.90, 0.95, 0.99$.

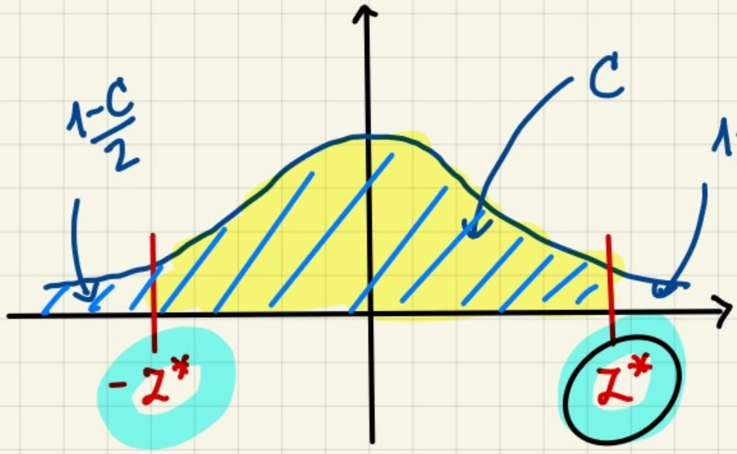
Look @

$$\mathbb{P}\left[\,|\overline{X} - \mu_X| < ?\,\right] = C$$

$$\boxed{\mathbb{P}\left[-? < \overline{X} - \mu_X < ?\right] = C} \quad (*)$$

$$\mathbb{P}\left[-\frac{?}{\frac{\sigma_X}{\sqrt{n}}} < \underbrace{\frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}}_{\overset{!!}{=}\ Z} < \frac{?}{\frac{\sigma_X}{\sqrt{n}}}\right] = C$$

$\sim N(0,1)$



$$C + \frac{1-C}{2} = \frac{1+C}{2}$$

$$z^* = \Phi^{-1}\left(\frac{1+C}{2}\right)$$

z^* is a CRITICAL VALUE such that

$$\mathbb{P}\left[-z^* < Z < z^*\right] = C$$

$$\Rightarrow \quad \frac{?}{\frac{\sigma_X}{\sqrt{n}}} = z^*$$

$$\boxed{? = z^* \cdot \frac{\sigma_X}{\sqrt{n}}}$$

$$\circledast \Rightarrow \quad \mathbb{P}\left[-z^* \cdot \frac{\sigma_X}{\sqrt{n}} < \bar{X} - \mu_X < z^* \cdot \frac{\sigma_X}{\sqrt{n}}\right] = C$$

$$\Leftrightarrow \quad \mathbb{P}\Big[\underbrace{-z^* \cdot \frac{\sigma_X}{\sqrt{n}} + \bar{X}} < \mu_X < \underbrace{z^* \cdot \frac{\sigma_X}{\sqrt{n}} + \bar{X}}\Big] = C$$

this is a random interval