Week 7. Central Limit Theorem

"Samples" of variable X are written $X_1,\ X_2, \dots,\ X_n$

"first measurement of X", "second measurement of X", etc…

(independent, identically distributed)

Observed values of sampling are lower-case: $x_1, x_2, \dots, x_n$

"Sample mean" of a set of samples is

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Observed sample mean is lower-case:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Example: Measuring heights X of 3 people yields sample values:

$$x_1 = 70\ inches$$
$$x_2 = 73\ inches$$
$$x_3 = 68\ inches$$

The sample mean is

$$\overline{x} = \frac{70 + 73 + 68}{3}\ inches$$

Clever idea: Sample mean $\overline{X}$ is *itself* a random variable!

(To measure $\overline{X}$, gather samples and compute average.)

- **Sample mean** values are *centered at* **population mean**: $\overline{x} \approx \mu_X$
- When n is big (more samples), it should be *very close* to μ_X!

Basic Facts about $\overline{X}$.

- $\overline{X}$ and X have the same expected value: $E\left[\,\overline{X}\,\right] = E[X] \ = \mu_X$
- $\overline{X}$ has std dev $\sigma_{\overline{X}}$ which gets small when n is big: $\sigma_{\overline{X}} = \dfrac{\sigma_X}{\sqrt{n}}$

Central Limit Theorem.

- $\overline{X}$ is *approximately* normal if n is "big"*

*(In practice, this means "$n > 30$")

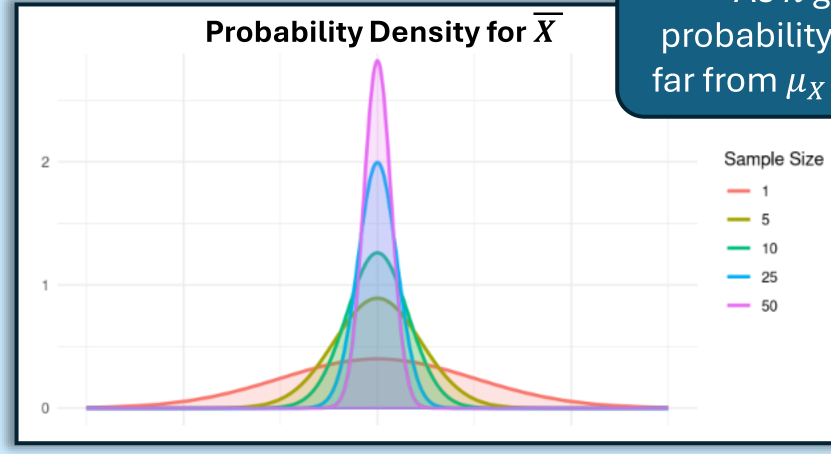
Recall: Normal distribution measures *"sums of lots of small effects"*…

That is **exactly** what a sample mean is! $\overline{X} = \frac{X_1}{n} + \frac{X_2}{n} + \cdots \frac{X_n}{n}$

This means we can compute ***z*-scores** and **probabilities** of $\overline{x}$ values (for each n)!

⇨ We already knew $\overline{x}$ *"should"* be *"close"* to μ_X, but now we can compute probabilities!

Week 7. z-Scores and Probability



As n gets big, the probability of measuring $\overline{x}$ far from μ_X gets **VERY LOW**

- Probability of $\overline{X}$ is centered at population mean μ_X.
- For bigger n, it is thinner and taller.

z-score of $\overline{X}$ value

$$z = \frac{\overline{x} - \mu_X}{\sigma_X / \sqrt{n}}$$

Recall. z-score of X value is $z = \frac{x - \mu}{\sigma}$

Example.

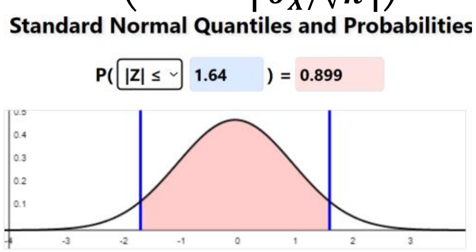
Suppose $\mu_X = 20$, $\sigma_X = 2$, and $\overline{x} = 22$.

- If $n = 1$, then $z = \frac{22-20}{2/\sqrt{1}} = \frac{2}{2} = 1$ "common"
- If $n = 4$, then $z = \frac{22-20}{2/\sqrt{4}} = \frac{2}{1} = 2$ "rare"
- If $n = 16$, then $z = \frac{22-20}{2/\sqrt{16}} = \frac{2}{1/2} = 4$ "very rare"

In the next chapters, we will apply these z-scores to compute a few different types of probability for sample mean values (for specific numbers of samples n).

Interior probability:

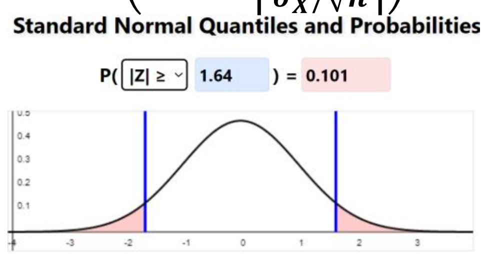
$$P\left(|Z| \le \left|\frac{\overline{x}-\mu_X}{\sigma_X/\sqrt{n}}\right|\right)$$



Confidence Intervals

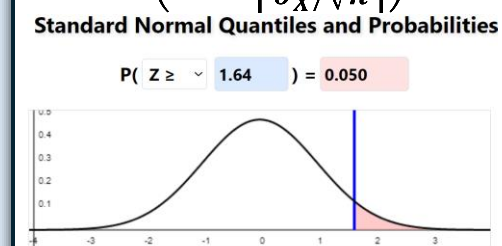
Two-tailed probability:

$$P\left(|Z| \ge \left|\frac{\overline{x}-\mu_X}{\sigma_X/\sqrt{n}}\right|\right)$$



One-tail probability:

$$P\left(Z \ge \left|\frac{\overline{x}-\mu_X}{\sigma_X/\sqrt{n}}\right|\right)$$



Hypothesis Testing