Unit 5 ♡♡

Normal Distribution

- Normal: symmetric, mounded in middle, bell curve, from a normal population or Sample size > 30.

★ Keep discrete graphs and continuous graphs separated. If discrete, add up probability values to find what's needed. If continuous with fractions, use z-score. Real world usually continuous.

Central limiting theorem

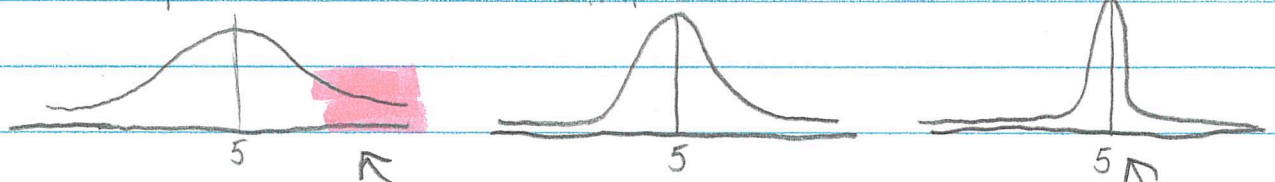
- Law of large #'s - More Samples run = closer to actual value. This means that eventually $\mu_{\bar{x}} = \mu$ (mean samples = true mean)

Essentially: $\bar{x} = \mu$ over repeated Samples (center samp. dist = pop center)

numerical → $\sigma_x = \frac{\sigma}{\sqrt{n}}$ over repeated samples because as you increase the sample size, the variance will decrease due to Central limiting theorem. (n = sample size)

Therefore... repeated samples = Same center, less variation so higher mounded data. (less spread out)

ex) 20 samples: 100 samples: 1000 samples:



Also note... less samples = larger tail ends
more samples = mounded in middle
more samples = closer to true mean

Categorical

$M\hat{p} = Mp$ Center population = Center samp dist.

$\sigma_s = \sqrt{\frac{p(1-p)}{n}}$ ← probability M, ← sample size

Conditions

numerical		Categorical	
random	Says	random	Says
Independent - small enough?	$n < .1N$	Independent	$n < .1N$ (sample size ↑, population ↑)
normal - big enough?	$n \geq 30$	normal $n \cdot p \geq 10$	and $n(1-p) \geq 10$

Bias check

- $M\hat{p} \pm Mp$. If more than 2 standard deviations away from parameter, the samples are likely biased. (If z-score of $M\hat{p}$ is greater than ± 2, it is likely biased.)

POINT ESTIMATOR – Statistic gathered from sample thats used to estimate parameter of population. Check for bias and conditions to see if statistic is good point estimator.

Sampling Distribution Differences

Numerical	Categorical
$M\hat{p}_1 - M\hat{p}_2 = p_1 - p_2$	$M\bar{x}_1 - M\bar{x}_2 = M_1 - M_2$
$\sigma_{\hat{p}_1} - \sigma_{\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$\sigma_{x1} - \sigma_{x2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
• Use to find differences in Means or SD from different groups.	• Use to find differences in proportions from different groups