C7 TEST TIPS!

Important Vocabulary

- **parameter**: a number that describes a population
- **statistic**: a number that describes a sample
- **population distribution**: of a variable describes the values of the variable for all individuals of a population
- **sampling distribution:** of a statistic describes the values of the statistic in all possible samples of the same size from the same population
- **distribution of sample data:** gives the values of the variable for all individuals in a particular sample
- **unbiased estimator:** does not consistently overestimate or underestimate the true value of the parameter, and the average of its estimates over many samples will be close to the true value.
- **variability:** of a statistic is described by the spread of its sampling distribution. Larger samples give smaller spread.
    - Taking a larger sample will reduce variability of a statistic, **but won't eliminate bias!!**
    - Larger samples will make a Normal curve taller and more narrow.
    - Smaller samples will make a Normal curve wider and less tall.
    - Read page 433.

When you see the language "check conditions" for a sampling distribution of a sample proportion

- **SRS**: determine if an SRS was taken or mention it was stated in the problem
- **Normality**: BINS / BITS - falls under the "Normality" heading because it's checking binomial and then Normal approximation of Binomial distribution.
- Then use the **large counts conditions:** $np \geq 10 \;\; n(1-p) \geq 10$. Put "stated" if it is stated in the problem that it is Normally or approximately Normally distributed.
- CLT can NEVER be used to check Normality with proportions!!!
- **10% condition:** check that $N \geq 10n$ where n is the sample size and N is the population size

When you see the language "check conditions" for a sampling distribution of a sample mean

- **SRS**: determine if an SRS was taken or mention it was stated in the problem
- **Normality**: use CLT if applicable, you must write the entire sentence "Because n ≥ 30 [plug in n], the sampling distribution is approx. Normal by CLT." Put "stated" if it is stated in the problem that it is Normally or approximately Normally distributed
- If the population distribution follows a Normal distribution, then so will the sampling distribution.
- **10% condition:** check that $N \geq 10n$ where n is the sample size and N is the population size

Mean and Standard Deviation of the Sampling Distribution of $\hat{p}$

- $\mu_{\hat{p}} = p$ since $\hat{p}$ is an unbiased estimator of p
- $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ for an SRS of size n, as long as the 10% condition is met

Mean and Standard Deviation of the Sampling Distribution of $\bar{x}$

- $\mu_{\bar{x}} = \mu$ since $\bar{x}$ is an unbiased estimator of μ
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ as long as the 10% condition is met

The Sampling Distribution of $\hat{p}$

- describes how the sample proportion $\hat{p}$ varies in **all** possible samples from the population

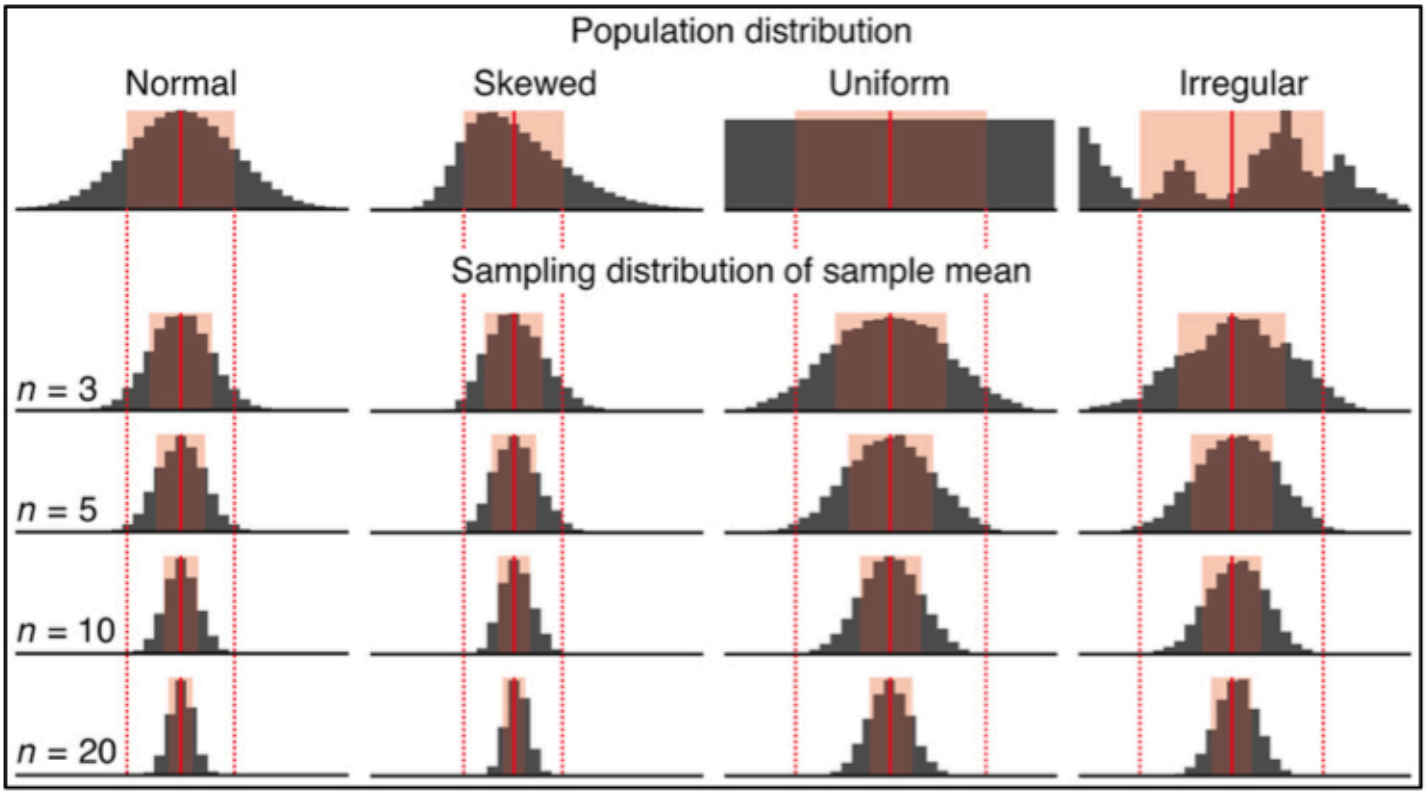
The Sampling Distribution of $\bar{x}$

- describes how the sample mean $\bar{x}$ varies in **all** possible samples of the same size from the population
- depends on sample size, not population size
- If n is small and we don't know the shape of the population distribution, then we don't know the shape of the sampling distribution

Central Limit Theorem

- States that as the sample size increases, the **distribution of the sample mean** will become more and more Normally distributed, regardless of the shape of the original distribution.
- If $n \geq 30$, the CLT tells use that the sampling distribution of $\bar{x}$ will be approximately Normal in most cases.
- CLT only refers to the SHAPE of the sampling distribution, not the center or spread.
- CLT can only be used for means, NOT PROPORTIONS.

Note the changes as sample size increases



https://mat117.wisconsin.edu/wp-content/uploads/2014/12/section7-1.png

z-scores

- Sometimes you will see probabilities in terms of z scores
- Remember z = (statistic - parameter) / (standard deviation of statistic)
- If we are finding a z-score from sampling distribution of $\bar{x}$, the formula becomes $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- If we are finding a z-score from sampling distribution of $\hat{p}$, the formula becomes $z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

When using a Normal Distribution

- Yes, always draw Normal curves!
- Include notation N(mean, standard deviation)
- Always write a probability statement: P(X < value), P(z > value), P($\hat{p}$ < value), etc…
- Shade and label graph
- You can use language like: $\bar{x}$ has approximately the N(45, 6.2) distribution
- Please DO NOT write calculator syntax for Normal probabilities, as it does not count for points on AP Statistics tests and the AP Exam. AP readers have said that they treat NormalCDF as invisible; they do not even look at it. The graph is what is required to earn work credit on Normal probability calculations.

You got this!