Central Limit Theorem (CLT)

Central Limit Theorem (CLT)

The **Central Limit Theorem (CLT)** is one of the most important concepts in statistics. It states that:

✅ **When we take many random samples from any population (no matter the original distribution), the sample means will follow a normal distribution if the sample size is large enough**.

✅ **The larger the sample size, the closer the distribution of the sample means will be to a normal distribution, regardless of the shape of the original population distribution**.

---

1. Central Limit Theorem Explained

Let's say we have a population with any distribution (it can be **uniform, skewed, exponential, or anything**). If we:

1. **Take repeated random samples of size n** from this population.
2. **Calculate the mean $\bar{X}$ of each sample**.
3. **Plot the distribution of these sample means**.

Then, **as n increases**, the distribution of these sample means will approach a **normal distribution** with:

- **Mean**: (same as the population mean).

  $\mu_{\bar{X}} = \mu$

- **Standard Deviation (Standard Error)**:

  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
  where
  σ is the population standard deviation, and n is the sample size.

---

2. Why is CLT Important?

🔷 **Allows us to use normal probability models** even when the population is not normal.

🔷 **Helps in hypothesis testing and confidence intervals**, as we can assume normality for large enough samples.

🔷 **Makes statistical inference possible**, as real-world data is often non-normal.

---

3. Example: Dice Rolling

🎲 Suppose we roll a **fair six-sided die**:

- The population distribution is **uniform** (1, 2, 3, 4, 5, 6 all equally likely).
- If we roll **one die**, the distribution is **uniform**, not normal.
- If we roll **two dice** and take the average, the distribution starts looking more **bell-shaped**.
- If we roll **30 dice** and take the average, the distribution of the means **closely follows a normal distribution**.

This is the **Central Limit Theorem in action**!

---

4. Conditions for CLT to Work

✅ **Sample Size is Large**: $n \geq 30$ is generally sufficient.

✅ **Random Sampling**: Samples should be independent and randomly selected.

✅ **Finite Variance**: The population should have a finite standard deviation.

---

5. Summary

Concept	Description
What CLT Says	The distribution of sample means approaches normal as sample size increases.
Applies To	Any population distribution (even skewed or non-normal).
Mean of Sample Means	$\mu_{\bar{X}} = \mu$(same as the population mean).
Standard Error (Spread)	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Minimum Sample Size	$n \geq 30$ is usually enough.

How To determine whether the Central Limit Theorem (CLT) should be used

To determine whether the **Central Limit Theorem (CLT)** should be used to solve a problem, you need to look for specific clues in the question. Here's how you can deduce that this problem is supposed to be solved using the CLT:

In CLT we are asked to approximate the probability that the sum of some object, also Approximate the probability that the sample mean of some object. both of these cases have different formulas.

Case	**Approximate Distribution**	**Standardization Formula**
Sum S_n	$N(n\mu, n\sigma^2)$	$Z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$
Sample Mean $\bar{X}$	$N(\mu, \frac{\sigma^2}{n})$	$Z = \frac{\bar{X} - \mu}{\sigma\sqrt{n}}$

Key Clues in the Problem

1. **Sample Mean is Involved**:
   - The problem asks about the probability that the **sample mean** lies within a certain range (e.g., $P\{124 < \bar{X} < 132\}$).
   - The CLT is specifically about the distribution of the **sample mean** when the sample size is large.
2. **Population Parameters are Given**:
   - The problem provides the **population mean** μ and **population standard deviation** σ.
   - The CLT uses these parameters to describe the distribution of the sample mean.
3. **Sample Size is Mentioned**:
   - The problem specifies different sample sizes $(n = 9, 25, 100)$.
   - The CLT applies when the sample size is "large enough" (typically $(n \geq 30)$), but it can also be used for smaller sample sizes if the

population distribution is not heavily skewed.

4. **Approximation is Requested**:
    - The problem asks for an **approximate probability**, which is a strong indicator that the CLT should be used. The CLT provides an approximation of the sampling distribution of the sample mean.

Why the CLT is Appropriate Here

- The problem involves calculating the probability that the **sample mean** lies between two values 124 and 132 .
- The population mean $\mu = 128$ and standard deviation $\sigma = 16$ are given.
- The sample sizes (n = 9, 25, 100) are provided, and the CLT can be applied even for smaller sample sizes if the population distribution is not heavily skewed.
- The problem asks for an **approximate probability**, which is exactly what the CLT provides.

When to Use the CLT

The CLT is used when:

1. You are dealing with the **sample mean** (not individual observations).
2. The sample size is "large enough" typically $n \geq 30$, or the population distribution is normal.
3. You are asked to find probabilities or percentiles related to the sample mean.

Conclusion

In this problem, the CLT is the correct tool because:

- The question involves the **sample mean**.
- The population mean and standard deviation are given.
- The sample sizes are provided, and the CLT can be applied even for smaller sample sizes.
- The problem asks for an **approximate probability**.
