Confidence Intervals and Their Connection to Hypothesis Testing

- The central limit theorem tells us how far $\bar{x}$ and μ are likely to be from each other.
- In hypothesis testing, we assume μ is fixed and ask how far $\bar{x}$ is from μ.
- In confidence intervals, we assume $\bar{x}$ is fixed and ask how far μ could reasonably be from $\bar{x}$.
- Because the variability is the same, these are just two sides of the same coin.

---

1. What is a Confidence Interval (CI)?

A confidence interval gives us a range of plausible values for the true population mean (μ) based on the sample mean ($\bar{x}$). It answers the question:

"What values of the population mean are consistent with the data we observed?"

2. Why Confidence Intervals Center Around $\bar{x}$

- The **sample mean ($\bar{x}$)** is our best guess for the true population mean (μ) based on the data.
- Since we don't know μ, we construct the confidence interval around $\bar{x}$, reflecting how much uncertainty we have due to sampling variability.

3. Variability of the Sample Mean

Thanks to the **central limit theorem (CLT)**, if we repeatedly sampled from the population, the sample mean ($\bar{x}$) would follow a normal distribution:

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \tag{1}$$

where:

- μ = population mean.
- $\frac{\sigma}{\sqrt{n}}$ = standard error of the mean.

This means we know how far apart μ and $\bar{x}$ are likely to be.

4. Confidence Interval Formula

For a confidence level of $C\%$ (e.g., 95%), the confidence interval is:

$$\text{CI} = \bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}} \tag{2}$$

where:

- z^*: critical value for the confidence level (from the standard normal table).
- $\frac{\sigma}{\sqrt{n}}$: standard error of the mean.

If the population standard deviation (σ) is unknown, we use the sample standard deviation (s) and the **t-distribution**:

$$\text{CI} = \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}} \tag{3}$$

5. Interpretation of Confidence Intervals

A 95% confidence interval means:

- **"If we repeated this sampling process many times, 95% of the intervals we construct would contain the true mean (μ)."**
- It does **not** mean there's a 95% chance that μ is in this particular interval. The true mean is fixed; it's the interval that varies.

These seem like similar ideas, but we need to be careful to state it in a way consistent with the population mean being non-random.

6. Connection to Hypothesis Testing

Confidence intervals are closely tied to hypothesis testing. Both involve the same fundamental question: **How far apart can the population mean (μ) and the sample mean ($\bar{x}$) be, given the variability in the data?**

In hypothesis testing:

- Assume $\mu = \mu_0$, where μ_0 is the null hypothesis.
- Use the sampling distribution centered at μ_0 to decide if $\bar{x}$ is consistent with μ_0.

In confidence intervals:

- Assume $\bar{x}$ is fixed (from the observed sample).
- Use the same variability to find all plausible values for μ.

The two approaches are mathematically equivalent:

- If μ_0 lies outside the confidence interval, we would reject H_0 in a hypothesis test with the same significance level.

7. Why Both Perspectives Work

There's symmetry in how the sample mean ($\bar{x}$) and population mean (μ) relate:

- The sampling distribution describes the variability of $\bar{x}$ around μ.
- This same distribution tells us how far μ could be from $\bar{x}$.

Whether we center the distribution on μ (for hypothesis testing) or $\bar{x}$ (for confidence intervals), the results are equivalent because the variability is the same.

8. Example

You take a sample of (n = 25) students and find:

- $\bar{x} = 65$ inches (mean height).
- $s = 3$ inches (sample standard deviation).

Construct a 95% confidence interval for the population mean (μ).

1. Find the standard error:

$$\text{SE} = \frac{s}{\sqrt{n}} = \frac{3}{\sqrt{25}} = 0.6 \tag{4}$$

2. Find t^* (from a t-table or code, with df = 24): $t^* = 2.064$.
3. Calculate the confidence interval:

$$\text{CI} = 65 \pm 2.064 \cdot 0.6 = [63.76, 66.24] \tag{5}$$

Interpretation: We are 95% confident that the true mean height of all students is between 63.76 and 66.24 inches.

9. Key Takeaways

1. Confidence intervals estimate a range for the true population mean, centered at $\bar{x}$.
2. The interval reflects sampling variability, using the standard error.
3. Confidence intervals and hypothesis tests ask the same fundamental question about how far apart μ and $\bar{x}$ can reasonably be.