CHAPTER 8: INTERVAL ESTIMATION

Confidence Intervals and the Margin of Error

I. The Concept of an Interval

A point estimate (like $\bar{x}$) is rarely exactly equal to the population parameter (μ). An **Interval Estimate** accounts for this by providing a range:

$$\text{Point Estimate} \pm \text{Margin of Error}$$

The **Confidence Level** (e.g., 90%, 95%, 99%) represents the probability that the interval will contain the true population parameter if we were to take many samples.

II. Interval Estimate of μ

How we calculate the interval depends on whether the population standard deviation (σ) is known.

1. σ Known (z-Distribution)

If σ is known, we use the standard normal distribution:

$$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Common z values: 95% CI $\rightarrow z = 1.96$; 99% CI $\rightarrow z = 2.576$.

2. σ Unknown (t-Distribution)

In practice, if we don't know μ, we rarely know σ. We use the sample standard deviation (s) and the t-**distribution**:

$$\bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

The t-Distribution:

- Depends on **Degrees of Freedom (df)**, where $df = n - 1$.

- As df increases, the t-distribution approaches the z-distribution.

- It has "thicker tails" than the normal distribution to account for the extra uncertainty of estimating σ.

III. Interval Estimate of p

For categorical data, the interval for the population proportion is:

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Note: We always use z for proportions, provided $n\bar{p} \geq 5$ and $n(1-\bar{p}) \geq 5$.

IV. Determining Sample Size

In business planning, we often ask: "How many people do I need to survey to be 95% sure my error is less than $10?"

1. For the Mean:

$$n = \frac{(z_{\alpha/2})^2\sigma^2}{E^2}$$

Where E is the desired **Margin of Error**.

2. For the Proportion:

$$n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2}$$

If no planning value p^* is available, use 0.50 (the "worst-case" scenario for spread).

V. Teacher's Strategy: The Relationship

To master this chapter, understand these three relationships:

- **Confidence Level ↑, Interval Width ↑:** To be *more* sure, your net must be wider.

- **Sample Size (n) ↑, Interval Width ↓:** More data leads to more precision (smaller error).

- **Variation (σ) ↑, Interval Width ↑:** More "noise" in the data makes for less certain estimates.

VI. Step-by-Step Example

Problem: A sample of $n = 25$ coffee shops shows an average daily revenue of $\bar{x} = \$800$ with $s = \$50$. Calculate a 95% CI.

Logic: 1. **Identify Model:** σ is unknown, so use t. 2. **Degrees of Freedom:** $25 - 1 = 24$. 3. **Find t-value:** Looking at a t-table for $df = 24$ and 0.025 in the tail, $t_{0.025} = 2.064$. 4. **Calculate Std Error:** $50/\sqrt{25} = 10$. 5. **Calculate Margin of Error:** $2.064 \times 10 = 20.64$. 6. **Interval:** $800 \pm 20.64 \rightarrow [\$\mathbf{779.36}, \$\mathbf{820.64}]$.

VII. Practice Set

1. You want a 95% CI for a proportion with a margin of error of 0.03. Using $p^* = 0.50$, what is the required n?

2. As degrees of freedom increase, what happens to the t-value for a specific confidence level?

3. A 90% CI is $[50, 70]$. What is the point estimate and the margin of error?

4. When should you use a z-score instead of a t-score for a population mean?

VIII. Answer Key

1. $n = \frac{(1.96)^2(0.5 \times 0.5)}{0.03^2} = \frac{0.9604}{0.0009} \approx \mathbf{1068}$. 2. It **decreases** (moves closer to the z-value). 3. **Estimate:** 60 (midpoint); **Error:** 10 (half-width). 4. Only when the population standard deviation (σ) is **known**.