LARGE SAMPLE CONFIDENCE INTERVALS

MATH 3342
Section 7.2

Previous Assumptions

Suppose μ is the parameter of interest:

1. The data is from a SRS. There are no non-sampling errors.
2. The variable of interest is exactly Normal distributed.
3. We don't know the population mean μ, but we **do** know the population standard deviation σ.

What if these assumptions are not valid?

Large Sample CI with σ Known

- Let $X_1,\ldots,X_n$ be a SRS from a population with mean μ and standard deviation σ.
- If *n* is large, then $\overline{X}$ is approximately Normally distributed.
- It follows that

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

is a large-sample confidence interval for μ with an approximate confidence level of $100(1-\alpha)\%$

What if σ is not known?

- If we replace σ with *s*, then

$$\frac{\overline{X}-\mu}{S/\sqrt{n}}$$

- has an approximately normal distribution for large *n*.

- It follows that a large sample CI with a confidence level of approximately 100(1-α)% is

$$\overline{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

When is this CI valid?

- $X_1,\ldots,X_n$ is a SRS from a population with unknown mean μ and unknown standard deviation σ.
- The population distribution can have *any shape*.
- $n > 40$
  - More strict than the CLT because both $\bar{X}$ and S are random.
  - For smaller samples, we must use a slightly different approach.
    - Section 7.3

The Traditional CI for the Proportion *p*

Sample Proportion

$$\hat{p} = \frac{\#\ successes}{n} = \frac{x}{n}$$

- The sampling distribution is approximately Normal (for large n)
- The mean is the true proportion, p.
- The standard deviation is estimated by $s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Traditional Large Sample Confidence Interval

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

The Score CI for *p*

Let $\tilde{p} = \dfrac{\hat{p} + z^2_{\alpha/2}/(2n)}{1 + z^2_{\alpha/2}/n}$.

Then a CI for p with a confidence level of approximately $100(1-\alpha)\%$ is:

$$\tilde{p} \pm z_{\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})/n + z^2_{\alpha/2}/(4n^2)}}{1 + z^2_{\alpha/2}/n}$$

Example

- 11 out of 40 randomly selected customers are willing to pay $100 for a software upgrade.
- Construct a 95% confidence interval for the proportion of all customers willing to pay $100.

$$\hat{p} = \frac{11}{40} = 0.275$$

$$\tilde{p} = \frac{\hat{p} + z^2_{\alpha/2}/(2n)}{1 + z^2_{\alpha/2}/n} = \frac{0.275 + 1.96^2/(80)}{1 + 1.96^2/40} = 0.2947$$

$$z_{\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})/n + z^2_{\alpha/2}/(4n^2)}}{1 + z^2_{\alpha/2}/n} = 1.96 \frac{\sqrt{0.275(0.725)/40 + 1.96^2/6400}}{1 + 1.96^2/40} = 0.1414$$

Large-Sample Confidence Bounds

- A large-sample **upper confidence bound** for μ is:

$$\mu < \bar{x} + z_{\alpha} \cdot \frac{s}{\sqrt{n}}$$

- A large-sample **lower confidence bound** for μ is:

$$\mu > \bar{x} - z_{\alpha} \cdot \frac{s}{\sqrt{n}}$$

- Also called **one-sided confidence intervals**.
- Can be calculated similarly for *p*.