# Confidence Intervals for the Population Mean

MATH 2300
Chapter 8

## Example

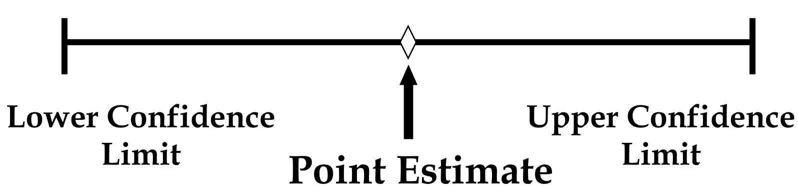
- A sample of 100 soda cans, from a population with soda volume being Normally distributed having $\sigma = 0.20$ oz, produced a sample mean equal to 12.09 oz.

- What is a reasonable **point estimate** for the population mean $\mu$?
- What is a likely range of values for the population mean $\mu$ of soda volume?

# Point Estimates

- A **point estimate** of a parameter is the value of a statistic that we are using to estimate the parameter.
- A statistic is called **unbiased** if its mean value is the true value of the parameter.
  - Otherwise, the statistic is **biased**.

# Confidence Intervals



# Confidence Intervals

- Provides a range of plausible values for a parameter.
  - Developed from a sample
  - For a given confidence level
- A **confidence level C** is a measure of the degree of reliability of the interval.
  - Gives the probability that the interval will capture the true parameter in repeated samples.
  - C is usually 90% or higher.

# Assumptions for Case I

Suppose $\mu$ is the parameter of interest:

1. The data is from a SRS. There are no non-sampling errors.
2. Either of the following is true:
   1. The variable is Normal distributed
   2. The sample size is large ($n \geq 30$)
3. We don't know the population mean $\mu$, but we **do** know the population standard deviation $\sigma$.
   - Not necessarily a practical assumption.

## Construction of a 95% CI

- We build our interval starting from the formula for a z-score and a probability statement about a standard normal RV.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \qquad P(-1.96 < Z < 1.96) = 0.95$$

## Construction of a 95% CI

$$\begin{aligned}
0.95 &= P(-1.96 < Z < 1.96) \\
&= P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < 1.96\right) \\
&= P\left(-1.96 \cdot \sigma / \sqrt{n} < \bar{X} - \mu < 1.96 \cdot \sigma / \sqrt{n}\right) \\
&= P\left(-\bar{X} - 1.96 \cdot \sigma / \sqrt{n} < -\mu < -\bar{X} + 1.96 \cdot \sigma / \sqrt{n}\right) \\
&= P\left(\bar{X} - 1.96 \cdot \sigma / \sqrt{n} < \mu < \bar{X} + 1.96 \cdot \sigma / \sqrt{n}\right)
\end{aligned}$$

# Construction of a 95% CI

$\left(\bar{X} - 1.96 \cdot \sigma/\sqrt{n}, \bar{X} + 1.96 \cdot \sigma/\sqrt{n}\right)$ is a Random Interval.

The width is not random, only the *center* is.

*Interpretation*: The probability that this random interval includes the true value of $\mu$ is 95%.

# 95% CI for $\mu$

Suppose we observe $X_1 = x_1, \ldots, X_n = x_n$ and compute $\bar{x}$ and use it in place of $\bar{X}$, then:

$\left(\bar{x} - 1.96 \cdot \sigma/\sqrt{n}, \bar{x} + 1.96 \cdot \sigma/\sqrt{n}\right)$ is a 95% CI for $\mu$.

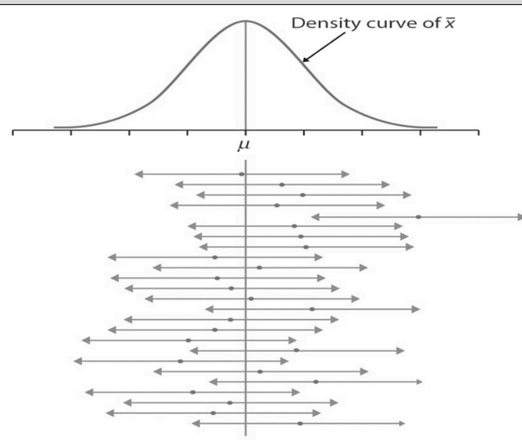
Alternatively,

$\bar{x} - 1.96 \cdot \sigma/\sqrt{n} < \mu < \bar{x} + 1.96 \cdot \sigma/\sqrt{n}$ with 95% confidence.

# Interpretation

- When we substitute in the observed sample mean, the interval is no longer random!
- Therefore:
  - The probability statement is not about any one observed interval
  - Rather, a statement about the **long-term relative frequency** if we repeatedly sample and compute intervals in this manner.

# Interpretation



# Example

A sample of 100 soda cans, from a population with soda volume being Normally distributed having σ = 0.20 oz, produced a sample mean equal to 12.09 oz. A 95% confidence interval would be:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

$$12.09 \pm 1.96 \frac{0.20}{\sqrt{100}}$$

$$12.09 \pm 0.039$$

**12.051 ounces** ———————— **12.129 ounces**

# 100(1- $\alpha$ )% Confidence Intervals

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

where: $z_{\alpha/2}$ = Standard Normal Critical Value

σ = Population standard deviation

n = Sample size

# Common Critical Values

| Confidence Level | 90% | 95% | 99% |
|---|---|---|---|
| $z_{\alpha/2}$ | 1.645 | 1.960 | 2.576 |

# Example

A sample of 100 soda cans, from a population with soda volume being Normally distributed having σ = 0.20 oz, produced a sample mean equal to 12.09 oz. A **99%** confidence interval would be:

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

$$12.09 \pm 2.576 \frac{0.20}{\sqrt{100}}$$

$$12.09 \pm 0.052$$

**12.051 ounces** ——— **12.129 ounces**

**12.038 ounces** ——— **12.142 ounces**

# How do we reduce width?

- Lower the confidence level. This results in a lower critical value $z_{\alpha/2}$.
- Increase the sample size $n$. This reduces the variability of the sampling distribution.

# Example: Impact of $n$

If instead of sample of 100 cans, suppose a sample of 400 cans, from a population with $\sigma = 0.20$, produced a sample mean equal to 12.09. A 95% confidence interval would be:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

$$12.09 \pm 1.96 \frac{0.20}{\sqrt{400}}$$

$$12.09 \pm 0.0196$$

**12.0704 ounces** ——— **n=400** ——— **12.1096 ounces**

**12.051 ounces** ——— **n=100** ——— **12.129 ounces**

# Margin of Error

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where: $z_{\alpha/2}$ = Standard Normal Critical Value

$\sigma$ = Population standard deviation

n = Sample size

# The Minimum $n$ for a Desired Width

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

where:

$z_{\alpha/2}$ = Critical value

$w$ = Desired width

$\sigma$ = Population standard deviation

# Always round up for *n* !!!

- Suppose the calculation says that 58.2 observations are needed.
- 58 observations would produce a slightly wider interval than is wanted.
- 59 observations would produce a slightly narrower interval than desired.
- Estimate would still then be within the desired margin of error with the desired level of confidence

# Example

The manager of the Georgia Timber Mill wishes to construct a 90% confidence interval within 0.50 inches in estimating the mean diameter of logs. Assume a population standard deviation of 4.8 inches.

$$n = \left(2 \cdot 1.645 \frac{4.8}{1}\right)^2 = 249.38 \boxed{\approx} 250$$

# Example

- A sample of 100 soda cans from a population with soda volume being Normally distributed produced a sample mean equal to 12.09 oz and a sample standard deviation *s* of 0.20 oz.
  - In this case, we can actually use our previous result
- **However**, what if we only had 10 soda cans instead?

# Assumptions for Case II

- The data is from a SRS.
- Observations from the population are either from:
  - A Normal distribution with **unknown** mean $\mu$ and **unknown** standard deviation $\sigma$.

    OR
  - A symmetric, single-peaked distribution with **unknown** mean $\mu$ and **unknown** standard deviation $\sigma$.
    - This assumption results in approximations

# Small-Sample Distribution

- Under these assumptions, for small or moderate *n*,

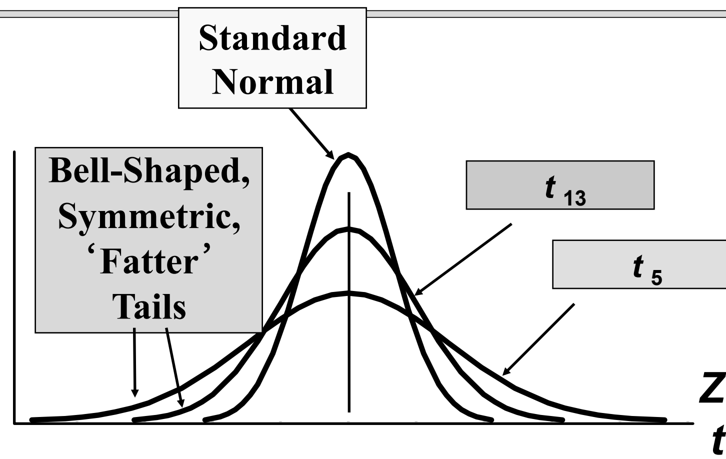
$$\frac{\bar{X}-\mu}{S/\sqrt{n}}$$

- Does NOT have a Normal distribution.
- It has a **t distribution** with *n* – *1****degrees of freedom***.
  - Also called Student's t distribution

# Student's t Distributions

- The t-distribution family has the following properties:
  - Bell-shaped and symmetric
  - Greater area in the tails than in the tails of the Normal
- The t-distribution approaches the normal distribution as the degrees of freedom increase.
- Let $t_v$ denote the t distribution with $v$ df.
- Critical values are provided in Table IV

# Convergence to the Normal Distribution




# One-Sample t Confidence Intervals

$$\bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

where:

$t_{\alpha/2,n-1}$ = Critical value for $t_{n-1}$

$\bar{x}$ = Sample mean

s = Sample standard deviation

$n$ = Sample size

# Example

- A sample of 10 soda cans, from a population with soda volume being Normally distributed having $s$ = 0.20 oz, produced a sample mean equal to 12.09 oz.

$$\bar{x} \pm t_{\alpha/2,9} \frac{s}{\sqrt{n}} \longrightarrow 12.09 \pm 2.262 \frac{0.20}{\sqrt{10}}$$

$$12.09 \pm 0.1431$$

**11.957 ounces** —— ***t*** **- CI** —— **12.233 ounces**

**11.966 ounces** —— **Z - CI** —— **12.214 ounces**

# Robustness

- An inference procedure is considered robust if the calculations required are **in**sensitive to violations in the assumptions of the procedure

# Robustness of the t-Interval

- The t interval is **not** robust to outliers
  - Because the t statistic is calculated using x-bar and s, neither of which is resistant to outliers.
- The t interval **is** robust to lack of Normality
  - But the data must come from a SRS still.
  - Due to the Central Limit Theorem

# General Guidelines for the t-Interval

| Sample Size | Shape of Parent Population Distribution |
|---|---|
| $n < 15$ | roughly bell-shaped, no outliers |
| $15 \le n < 40$ | roughly symmetric, no outliers |