
Lecture 12: Confidence Interval (II)

STAT 515 — Statistical Methods I

Dr. Yen-Yi Ho

Department of Statistics, University of South Carolina

1. Confidence interval estimation for a single population mean
 - Two scenarios
2. The t distribution

Theoretical Background: σ^2 Known

- **Scenario A:** sampling from a **normally distributed** population with the population variance **known** (σ^2 known).
- **Note:** in practice this is almost never the case.

Confidence Interval for μ (σ Known)

$$\bar{x} \pm z_{\alpha/2} \cdot \sigma_{\bar{x}}, \quad \text{where} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$\sigma_{\bar{x}}$ is the **standard error** of the sample mean.

Theory tells us that

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1),$$

and therefore the interval

$$\bar{x} \pm z_{\alpha/2} \cdot \sigma_{\bar{x}}$$

covers μ with probability 0.95 when $z_{\alpha/2} = 1.96$.

The practical problem

In real studies σ^2 is **unknown**.

Practical Application: σ Is Never Known

Sampling from a **normally distributed** population with population variance unknown (but s^2 computed from the sample).

Confidence Interval for μ (σ Unknown)

$$\bar{x} \pm t_{(\alpha/2, n-1)} \cdot s_{\bar{x}}, \quad \text{where} \quad s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

and $t_{\alpha/2}$ has $n - 1$ **degrees of freedom**.

Sampling from a population that is **not normally distributed** (or whose shape is unknown) but the **sample size is large**.

$$\bar{x} \pm t_{(\alpha/2, n-1)} \cdot s_{\bar{x}} \quad \text{if } \sigma \text{ is unknown}$$

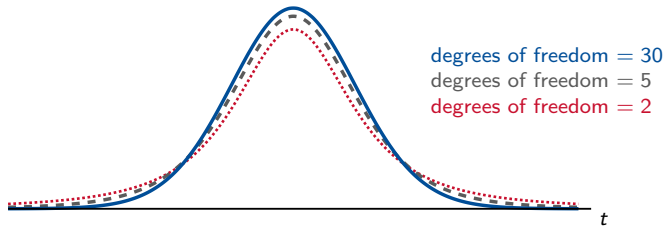
Large-sample behaviour

$t \longrightarrow z$ for large n : the t critical value converges to the normal critical value as the sample size grows.

1. Confidence interval estimation for a single population mean
2. The t distribution

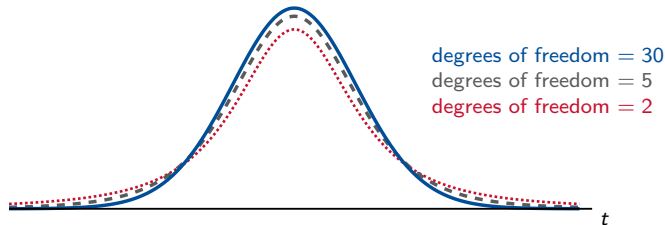
2. The t Distribution

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$



2. The t Distribution

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$



- The t -distribution is the “fatter, flatter cousin” of the normal: t -distribution is uniquely defined by **degrees of freedom**

2.1 Properties of the t Distribution

- Mean = median = mode = 0.
- **Symmetrical** about the mean.
- The value of t ranges from $-\infty$ to $+\infty$.
- A **family** of distributions determined by $n - 1$, the **degrees of freedom**.
- The t distribution approaches the **normal distribution** as $n - 1 \rightarrow \infty$.

Exercise 1: A Pollution Study

Problem 3

In a pollution study, a random sample of eight air specimens contained an average of 2.26 microgram/ m^3 of suspended matter with a standard deviation of .56 microgram/ m^3 . Find a 99% confidence interval for the mean amount of suspended matter in a 1 m^3 air sample in this area. (Assume the amount of suspended matter is normally distributed.)

Given: $\bar{x} = 2.26$, $n = 8$, $s = .56$, $1 - \alpha = 99\% \Rightarrow \alpha = .01 \Rightarrow \frac{\alpha}{2} = .005$.

Exercise 1: The Computation

Define the parameter

μ : the true mean amount of suspended matter in a 1 m^3 air sample in this area.

A 99% CI for μ

$$\bar{x} \pm t_{\alpha/2, df=n-1} \left(\frac{s}{\sqrt{n}} \right) = 2.26 \pm 3.499 \left(\frac{.56}{\sqrt{8}} \right) = 2.26 \pm .693 = (1.567, 2.953)$$

with $t_{.005, df=7} = 3.499$.

Exercise 1: Interpretation

Interpretation. We are 99% confident that the true mean amount of suspended matter is between 1.567 and 2.953 microgram/ m^3 .

- **Follow-up question:** what would a 95% CI for μ look like?
- Only the multiplier changes: use $t_{.025, df=7}$ in place of $t_{.005, df=7} = 3.499$.
- Since $t_{.025, df=7} < t_{.005, df=7}$, the 95% interval is **narrower** than the 99% interval.

Exercise 2: Small-Sample CI for μ

Exercise 1

The following random sample was selected from a normal distribution:

4, 6, 3, 5, 9, 3

- (a) Construct a 90% confidence interval for μ .
- (b) Construct a 95% confidence interval for μ .
- (c) Construct a 99% confidence interval for μ .

Setup: $n = 6$, $\bar{x} = 5$, $s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{26}{5} = 5.2$, $s = \sqrt{5.2} = 2.280$,
 $df = 5$.

Exercise 2: Solution

The standard error is $\frac{s}{\sqrt{n}} = \frac{2.280}{\sqrt{6}} = 0.931$.

(a) $1 - \alpha = .90 \Rightarrow \alpha/2 = .05$, $t_{.05,5} = 2.015$:

$$5 \pm 2.015(0.931) = 5 \pm 1.876 = (3.124, 6.876)$$

(b) $1 - \alpha = .95 \Rightarrow \alpha/2 = .025$, $t_{.025,5} = 2.571$:

$$5 \pm 2.571(0.931) = 5 \pm 2.393 = (2.607, 7.393)$$

(c) $1 - \alpha = .99 \Rightarrow \alpha/2 = .005$, $t_{.005,5} = 4.032$:

$$5 \pm 4.032(0.931) = 5 \pm 3.754 = (1.246, 8.754)$$

Key takeaway

The higher the confidence level, the **wider** the interval.