ISE 315: Engineering Statistics

Lecture 6: Statistical Intervals for a Single Sample (Part 2)

Instructor: Mansur M. Arief, PhD
Industrial and Systems Engineering, KFUPM

Office: 22-219 — Email: mansur.arief@kfupm.edu.sa

Based on Montgomery & Runger, Applied Statistics and Probability for Engineers, 6th Ed.

Lecture 6

Statistical Intervals for a Single Sample (Part 2)

Lecture 6 Outline (Textbook Chapter 3)

- Announcements:
  - HW #2 due next week
  - Major Exam 1 in Week 8 (Textbook Chapters 2–5)
  - Quiz #2 next Wednesday (up to 5 questions about Chapter 3)
- Z Table and Problem Examples
- CI for Mean (σ^2 unknown), Variance, and Proportion

How to Read the Z-Table (Table B.1 in the textbook appendix)

- The Z-table gives $\Phi(z) = P(Z \leq z)$ for $Z \sim N(0, 1)$

How to Read the Z-Table (Table B.1 in the textbook appendix)

- The Z-table gives $\Phi(z) = P(Z \leq z)$ for $Z \sim N(0, 1)$
- Rows: first digit and first decimal of z (e.g., 1.9)

How to Read the Z-Table (Table B.1 in the textbook appendix)

- The Z-table gives $\Phi(z) = P(Z \leq z)$ for $Z \sim N(0, 1)$
- Rows: first digit and first decimal of z (e.g., 1.9)
- Columns: second decimal of z (e.g., 0.06)

How to Read the Z-Table (Table B.1 in the textbook appendix)

- The Z-table gives $\Phi(z) = P(Z \leq z)$ for $Z \sim N(0, 1)$
- Rows: first digit and first decimal of z (e.g., 1.9)
- Columns: second decimal of z (e.g., 0.06)

z	**0.00**	**0.01**	**0.02**	$\cdots$	**0.05**	**0.06**	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
1.6	0.9452	0.9463	0.9474	$\cdots$	0.9505	0.9515	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
1.9	0.9713	0.9719	0.9726	$\cdots$	0.9744	0.9750	$\cdots$
2.0	0.9772	0.9778	0.9783	$\cdots$	0.9798	0.9803	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	

How to Read the Z-Table (Table B.1 in the textbook appendix)

- The Z-table gives $\Phi(z) = P(Z \leq z)$ for $Z \sim N(0, 1)$
- Rows: first digit and first decimal of z (e.g., 1.9)
- Columns: second decimal of z (e.g., 0.06)

z	**0.00**	**0.01**	**0.02**	$\cdots$	**0.05**	**0.06**	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
1.6	0.9452	0.9463	0.9474	$\cdots$	0.9505	0.9515	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
1.9	0.9713	0.9719	0.9726	$\cdots$	0.9744	0.9750	$\cdots$
2.0	0.9772	0.9778	0.9783	$\cdots$	0.9798	0.9803	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	

Example: $\Phi(1.96) = P(Z \leq 1.96) = 0.9750$

How to Find Z Critical Value (Inverse)

Given: A probability p, find z such that $P(Z \leq z) = p$

How to Find Z Critical Value (Inverse)

Given: A probability p, find z such that $P(Z \leq z) = p$

Method: Search *inside* the table for p, then read row $+$ column headers

How to Find Z Critical Value (Inverse)

Given: A probability p, find z such that $P(Z \leq z) = p$

Method: Search *inside* the table for p, then read row + column headers

Example: Find z such that $P(Z \leq z) = 0.975$

How to Find Z Critical Value (Inverse)

Given: A probability p, find z such that $P(Z \leq z) = p$

Method: Search *inside* the table for p, then read row + column headers

Example: Find z such that $P(Z \leq z) = 0.975$

z	**0.00**	$\cdots$	**0.05**	**0.06**	**0.07**	$\cdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	
1.8	0.9641	$\cdots$	0.9678	0.9686	0.9693	$\cdots$
1.9	0.9713	$\cdots$	0.9744	**0.9750**	0.9756	$\cdots$
2.0	0.9772	$\cdots$	0.9798	0.9803	0.9808	$\cdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	

How to Find Z Critical Value (Inverse)

Given: A probability p, find z such that $P(Z \leq z) = p$

Method: Search *inside* the table for p, then read row + column headers

Example: Find z such that $P(Z \leq z) = 0.975$

z	**0.00**	$\cdots$	**0.05**	**0.06**	**0.07**	$\cdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	
1.8	0.9641	$\cdots$	0.9678	0.9686	0.9693	$\cdots$
1.9	0.9713	$\cdots$	0.9744	**0.9750**	0.9756	$\cdots$
2.0	0.9772	$\cdots$	0.9798	0.9803	0.9808	$\cdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	

How to Find Z Critical Value (Inverse)

Given: A probability p, find z such that $P(Z \leq z) = p$

Method: Search *inside* the table for p, then read row + column headers

Example: Find z such that $P(Z \leq z) = 0.975$

z	**0.00**	$\cdots$	**0.05**	**0.06**	**0.07**	$\cdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	
1.8	0.9641	$\cdots$	0.9678	0.9686	0.9693	$\cdots$
1.9	0.9713	$\cdots$	0.9744	**0.9750**	0.9756	$\cdots$
2.0	0.9772	$\cdots$	0.9798	0.9803	0.9808	$\cdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	

Common Critical Values (Table 3.1 in the textbook)

Conf. Level	α	**Two-sided (CI)**		**One-sided**	
		$\Phi(z_{\alpha/2})$	$z_{\alpha/2}$	$\Phi(z_\alpha)$	z_α
90%	0.10	0.95	1.645	0.90	1.282
95%	0.05	0.975	1.960	0.95	1.645
99%	0.01	0.995	2.576	0.99	2.326

A logistics analyst measures container dwell time at a port. A random sample of $n = 36$ containers gives sample mean $\bar{x} = 3.750$ days. Historical data shows $\sigma = 0.90$ days. Construct a 95% CI for the true mean dwell time.

Example 1 (Review): Port Container Dwell Time

A logistics analyst measures container dwell time at a port. A random sample of $n = 36$ containers gives sample mean $\bar{x} = 3.750$ days. Historical data shows $\sigma = 0.90$ days. Construct a 95% CI for the true mean dwell time.

Example 1 (Review): Port Container Dwell Time

Given:

A logistics analyst measures container dwell time at a port. A random sample of $n = 36$ containers gives sample mean $\bar{x} = 3.750$ days. Historical data shows $\sigma = 0.90$ days. Construct a 95% CI for the true mean dwell time.

Example 1 (Review): Port Container Dwell Time

Given:

- $n = 36$, $\bar{x} = 3.750$ days, $\sigma = 0.90$ days

A logistics analyst measures container dwell time at a port. A random sample of $n = 36$ containers gives sample mean $\bar{x} = 3.750$ days. Historical data shows $\sigma = 0.90$ days. Construct a 95% CI for the true mean dwell time.

Example 1 (Review): Port Container Dwell Time

Given:

- $n = 36$, $\bar{x} = 3.750$ days, $\sigma = 0.90$ days
- Confidence level: 95% $\Rightarrow \alpha = 0.05 \Rightarrow z_{0.025} = 1.96$

A logistics analyst measures container dwell time at a port. A random sample of $n = 36$ containers gives sample mean $\bar{x} = 3.750$ days. Historical data shows $\sigma = 0.90$ days. Construct a 95% CI for the true mean dwell time.

Example 1 (Review): Port Container Dwell Time

Given:

- $n = 36$, $\bar{x} = 3.750$ days, $\sigma = 0.90$ days
- Confidence level: 95% $\Rightarrow \alpha = 0.05 \Rightarrow z_{0.025} = 1.96$

Plan: Two-sided z-interval, (3.1) via Procedure 3.1

Step 1: Calculate standard error

Step 1: Calculate standard error

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}} = \frac{0.90}{\sqrt{36}} = \frac{0.90}{6} = 0.15$$

Example 1: Solution

Step 1: Calculate standard error

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{0.90}{\sqrt{36}} = \frac{0.90}{6} = 0.15$$

Step 2: Calculate margin of error (3.3)

Example 1: Solution

Step 1: Calculate standard error

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}} = \frac{0.90}{\sqrt{36}} = \frac{0.90}{6} = 0.15$$

Step 2: Calculate margin of error (3.3)

$$E = z_{\alpha/2} \times \mathrm{SE} = 1.96 \times 0.15 = 0.294$$

Example 1: Solution

Step 1: Calculate standard error

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{0.90}{\sqrt{36}} = \frac{0.90}{6} = 0.15$$

Step 2: Calculate margin of error (3.3)

$$E = z_{\alpha/2} \times \text{SE} = 1.96 \times 0.15 = 0.294$$

Step 3: Construct the interval (3.1)

Example 1: Solution

Step 1: Calculate standard error

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{0.90}{\sqrt{36}} = \frac{0.90}{6} = 0.15$$

Step 2: Calculate margin of error (3.3)

$$E = z_{\alpha/2} \times \text{SE} = 1.96 \times 0.15 = 0.294$$

Step 3: Construct the interval (3.1)

$$\text{CI: } \bar{x} \pm E = 3.750 \pm 0.294$$

Step 1: Calculate standard error

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{0.90}{\sqrt{36}} = \frac{0.90}{6} = 0.15$$

Step 2: Calculate margin of error (3.3)

$$E = z_{\alpha/2} \times \text{SE} = 1.96 \times 0.15 = 0.294$$

Step 3: Construct the interval (3.1)

$$\text{CI: } \bar{x} \pm E = 3.750 \pm 0.294 = \boxed{(3.456,\ 4.044) \text{ days}}$$

Your startup wants to estimate the mean onboarding time of new users for its app. A pilot shows $\sigma = 12$ minutes. How many users must you observe so the margin of error is at most 3 minutes with 95% confidence?

Example 2 (Review): Sample Size Determination

Your startup wants to estimate the mean onboarding time of new users for its app. A pilot shows $\sigma = 12$ minutes. How many users must you observe so the margin of error is at most 3 minutes with 95% confidence?

Example 2 (Review): Sample Size Determination

Given:

Your startup wants to estimate the mean onboarding time of new users for its app. A pilot shows $\sigma = 12$ minutes. How many users must you observe so the margin of error is at most 3 minutes with 95% confidence?

Example 2 (Review): Sample Size Determination

Given:

- $\sigma = 12$ minutes, $E = 3$ minutes

Your startup wants to estimate the mean onboarding time of new users for its app. A pilot shows $\sigma = 12$ minutes. How many users must you observe so the margin of error is at most 3 minutes with 95% confidence?

Example 2 (Review): Sample Size Determination

Given:

- $\sigma = 12$ minutes, $E = 3$ minutes
- 95% confidence $\Rightarrow z_{0.025} = 1.96$

Your startup wants to estimate the mean onboarding time of new users for its app. A pilot shows $\sigma = 12$ minutes. How many users must you observe so the margin of error is at most 3 minutes with 95% confidence?

Example 2 (Review): Sample Size Determination

Given:

- $\sigma = 12$ minutes, $E = 3$ minutes
- 95% confidence $\Rightarrow z_{0.025} = 1.96$

Solution: apply (3.4)

Your startup wants to estimate the mean onboarding time of new users for its app. A pilot shows $\sigma = 12$ minutes. How many users must you observe so the margin of error is at most 3 minutes with 95% confidence?

Example 2 (Review): Sample Size Determination

Given:

- $\sigma = 12$ minutes, $E = 3$ minutes
- 95% confidence $\Rightarrow z_{0.025} = 1.96$

Solution: apply (3.4)

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 12}{3}\right)^2$$

Your startup wants to estimate the mean onboarding time of new users for its app. A pilot shows $\sigma = 12$ minutes. How many users must you observe so the margin of error is at most 3 minutes with 95% confidence?

Example 2 (Review): Sample Size Determination

Given:

- $\sigma = 12$ minutes, $E = 3$ minutes
- 95% confidence $\Rightarrow z_{0.025} = 1.96$

Solution: apply (3.4)

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 12}{3}\right)^2$$

$$= \left(\frac{23.52}{3}\right)^2 = (7.84)^2 = 61.47 \approx 62 \text{ users (round up!)}$$

An AV safety team measures the pedestrian-detection distance of a perception model in $n = 16$ test runs. Sample mean $\bar{x} = 185.0$ m, known $\sigma = 10$ m. Find a 95% lower confidence bound for the true mean detection distance.

Example 3 (Review): One-Sided Confidence Bound

An AV safety team measures the pedestrian-detection distance of a perception model in $n = 16$ test runs. Sample mean $\bar{x} = 185.0$ m, known $\sigma = 10$ m. Find a 95% lower confidence bound for the true mean detection distance.

Example 3 (Review): One-Sided Confidence Bound

Given:

An AV safety team measures the pedestrian-detection distance of a perception model in $n = 16$ test runs. Sample mean $\bar{x} = 185.0$ m, known $\sigma = 10$ m. Find a 95% lower confidence bound for the true mean detection distance.

Example 3 (Review): One-Sided Confidence Bound

Given:

- $n = 16$, $\bar{x} = 185.0$ m, $\sigma = 10$ m

An AV safety team measures the pedestrian-detection distance of a perception model in $n = 16$ test runs. Sample mean $\bar{x} = 185.0$ m, known $\sigma = 10$ m. Find a 95% lower confidence bound for the true mean detection distance.

Example 3 (Review): One-Sided Confidence Bound

Given:

- $n = 16$, $\bar{x} = 185.0$ m, $\sigma = 10$ m
- 95% one-sided $\Rightarrow z_{0.05} = 1.645$ (note: z_α, not $z_{\alpha/2}$!)

An AV safety team measures the pedestrian-detection distance of a perception model in $n = 16$ test runs. Sample mean $\bar{x} = 185.0$ m, known $\sigma = 10$ m. Find a 95% lower confidence bound for the true mean detection distance.

Example 3 (Review): One-Sided Confidence Bound

Given:

- $n = 16$, $\bar{x} = 185.0$ m, $\sigma = 10$ m
- 95% one-sided $\Rightarrow z_{0.05} = 1.645$ (note: z_α, not $z_{\alpha/2}$!)

Solution: apply the lower bound in (3.2)

An AV safety team measures the pedestrian-detection distance of a perception model in $n = 16$ test runs. Sample mean $\bar{x} = 185.0$ m, known $\sigma = 10$ m. Find a 95% lower confidence bound for the true mean detection distance.

Example 3 (Review): One-Sided Confidence Bound

Given:

- $n = 16$, $\bar{x} = 185.0$ m, $\sigma = 10$ m
- 95% one-sided $\Rightarrow z_{0.05} = 1.645$ (note: z_α, not $z_{\alpha/2}$!)

Solution: apply the lower bound in (3.2)

$$\mu \geq \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} = 185.0 - 1.645 \times \frac{10}{\sqrt{16}}$$

An AV safety team measures the pedestrian-detection distance of a perception model in $n = 16$ test runs. Sample mean $\bar{x} = 185.0$ m, known $\sigma = 10$ m. Find a 95% lower confidence bound for the true mean detection distance.

Example 3 (Review): One-Sided Confidence Bound

Given:

- $n = 16$, $\bar{x} = 185.0$ m, $\sigma = 10$ m
- 95% one-sided $\Rightarrow z_{0.05} = 1.645$ (note: z_α, not $z_{\alpha/2}$!)

Solution: apply the lower bound in (3.2)

$$\mu \geq \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} = 185.0 - 1.645 \times \frac{10}{\sqrt{16}}$$

$$= 185.0 - 1.645 \times 2.5 = 185.0 - 4.113 = \boxed{180.9 \text{ m}}$$

Key Terminologies (Part 2)

1. **t-Distribution:**
   - Used when σ is unknown
   - Has parameter $\nu =$ **degrees of freedom** (df)

Key Terminologies (Part 2)

1. **t-Distribution:**
   - Used when σ is unknown
   - Has parameter $\nu =$ **degrees of freedom** (df)
   - For sample mean: $\nu = n - 1$

Key Terminologies (Part 2)

1. **t-Distribution:**
   - Used when σ is unknown
   - Has parameter $\nu =$ **degrees of freedom** (df)
   - For sample mean: $\nu = n - 1$
   - Heavier tails than Normal

Key Terminologies (Part 2)

1. **t-Distribution:**
   - Used when σ is unknown
   - Has parameter $\nu =$ **degrees of freedom** (df)
   - For sample mean: $\nu = n - 1$
   - Heavier tails than Normal

2. **χ^2 (Chi-Square) Distribution:**
   - Used for variance σ^2
   - Only takes positive values
   - Skewed right (not symmetric!)

Key Terminologies (Part 2)

1. **t-Distribution:**
   - Used when σ is unknown
   - Has parameter ν = **degrees of freedom** (df)
   - For sample mean: $\nu = n - 1$
   - Heavier tails than Normal

2. **χ^2 (Chi-Square) Distribution:**
   - Used for variance σ^2
   - Only takes positive values
   - Skewed right (not symmetric!)

3. **Degrees of Freedom (df):**
   - Number of independent pieces of information
   - For n samples estimating μ: df $= n - 1$
   - Written as ν ("nu")

Key Terminologies (Part 2)

1. **t-Distribution:**
   - Used when σ is unknown
   - Has parameter $\nu =$ **degrees of freedom** (df)
   - For sample mean: $\nu = n - 1$
   - Heavier tails than Normal

2. **χ^2 (Chi-Square) Distribution:**
   - Used for variance σ^2
   - Only takes positive values
   - Skewed right (not symmetric!)

3. **Degrees of Freedom (df):**
   - Number of independent pieces of information
   - For n samples estimating μ: df $= n - 1$
   - Written as ν ("nu")

4. **Sample Proportion:**
   - $\hat{p} = X/n$ where $X =$ successes
   - Estimates proportion p
   - Uses Normal approximation (if some conditions are met)

Key Terminologies

5. Critical Values for Different Distributions

- $z_{\alpha/2}$: Critical value from standard Normal $N(0, 1)$
  - Used when σ is known or for proportions

Key Terminologies

5. Critical Values for Different Distributions

- $z_{\alpha/2}$: Critical value from standard Normal $N(0,1)$
  - Used when σ is known or for proportions
- $t_{\alpha/2,n-1}$: Critical value from t-distribution with $n-1$ df
  - Used when σ is unknown
  - Larger than corresponding z value

Key Terminologies

5. Critical Values for Different Distributions

- $z_{\alpha/2}$: Critical value from standard Normal $N(0, 1)$
  - Used when σ is known or for proportions
- $t_{\alpha/2,n-1}$: Critical value from t-distribution with $n-1$ df
  - Used when σ is unknown
  - Larger than corresponding z value
- $\chi^2_{\alpha/2,n-1}$ **and** $\chi^2_{1-\alpha/2,n-1}$: Critical values from χ^2-distribution
  - Need two values because χ^2 is not symmetric
  - Used for variance/standard deviation

Key Terminologies

5. Critical Values for Different Distributions

- $z_{\alpha/2}$: Critical value from standard Normal $N(0, 1)$
  - Used when σ is known or for proportions
- $t_{\alpha/2,n-1}$: Critical value from t-distribution with $n-1$ df
  - Used when σ is unknown
  - Larger than corresponding z value
- $\chi^2_{\alpha/2,n-1}$ **and** $\chi^2_{1-\alpha/2,n-1}$: Critical values from χ^2-distribution
  - Need two values because χ^2 is not symmetric
  - Used for variance/standard deviation

As df $\to \infty$: $t \to z$ and $\chi^2/df \to 1$

Key Terminologies

6. **Large Sample Conditions**
   - For t-interval: t approximates z when $n \geq 30$
     - But always use t when σ is unknown!

Key Terminologies

6. Large Sample Conditions

- For t-interval: t approximates z when $n \geq 30$
  - But always use t when σ is unknown!
- For proportion CI: Normal approximation is valid when:

$$n\hat{p} \geq 10 \quad \text{AND} \quad n(1-\hat{p}) \geq 10$$

  - At least 10 successes AND 10 failures

Key Terminologies

6. **Large Sample Conditions**
   - For t-interval: t approximates z when $n \geq 30$
     - But always use t when σ is unknown!
   - For proportion CI: Normal approximation is valid when:

$$n\hat{p} \geq 10 \quad \text{AND} \quad n(1-\hat{p}) \geq 10$$

     - At least 10 successes AND 10 failures
   - For χ^2-interval: **Requires Normal population**
     - Very sensitive to non-Normality
     - Check Normality assumption carefully

Key Terminologies

6. Large Sample Conditions

- For t-interval: t approximates z when $n \geq 30$
  - But always use t when σ is unknown!
- For proportion CI: Normal approximation is valid when:

$$n\hat{p} \geq 10 \quad \text{AND} \quad n(1-\hat{p}) \geq 10$$

  - At least 10 successes AND 10 failures
- For χ^2-interval: **Requires Normal population**
  - Very sensitive to non-Normality
  - Check Normality assumption carefully

When in doubt about assumptions, use larger sample sizes!

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
    - S is a random variable (varies from sample to sample)

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
  - S is a random variable (varies from sample to sample)
  - This adds extra uncertainty!

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
  - S is a random variable (varies from sample to sample)
  - This adds extra uncertainty!
- The statistic $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ does NOT follow $N(0, 1)$

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
    - S is a random variable (varies from sample to sample)
    - This adds extra uncertainty!
- The statistic $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ does NOT follow $N(0, 1)$
    - It follows the t-distribution with $n-1$ degrees of freedom

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
    - S is a random variable (varies from sample to sample)
    - This adds extra uncertainty!
- The statistic $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ does NOT follow $N(0,1)$
    - It follows the t-distribution with $n-1$ degrees of freedom
- Properties of t-distribution compared to Normal:

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
    - S is a random variable (varies from sample to sample)
    - This adds extra uncertainty!
- The statistic $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ does NOT follow $N(0,1)$
    - It follows the t-distribution with $n-1$ degrees of freedom
- Properties of t-distribution compared to Normal:
    - Also symmetric and bell-shaped

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
  - S is a random variable (varies from sample to sample)
  - This adds extra uncertainty!
- The statistic $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ does NOT follow $N(0,1)$
  - It follows the t-distribution with $n-1$ degrees of freedom
- Properties of t-distribution compared to Normal:
  - Also symmetric and bell-shaped
  - **Heavier tails** $\Rightarrow$ larger critical values

Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
  - S is a random variable (varies from sample to sample)
  - This adds extra uncertainty!
- The statistic $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ does NOT follow $N(0,1)$
  - It follows the t-distribution with $n-1$ degrees of freedom
- Properties of t-distribution compared to Normal:
  - Also symmetric and bell-shaped
  - **Heavier tails** $\Rightarrow$ larger critical values
  - t critical values $>$ z critical values

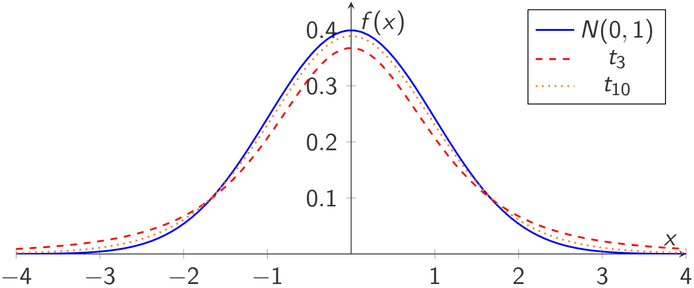
Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
  - S is a random variable (varies from sample to sample)
  - This adds extra uncertainty!
- The statistic $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ does NOT follow $N(0, 1)$
  - It follows the t-distribution with $n-1$ degrees of freedom
- Properties of t-distribution compared to Normal:
  - Also symmetric and bell-shaped
  - **Heavier tails** $\Rightarrow$ larger critical values
  - t critical values $>$ z critical values
  - Results in **wider** confidence intervals

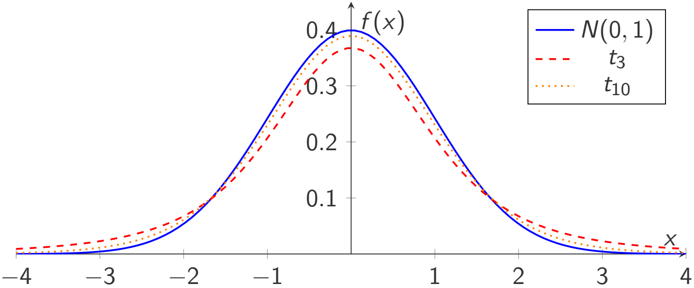
Notes about t vs Normal Distribution

- Why can't we just replace σ with s in the z-formula?
  - S is a random variable (varies from sample to sample)
  - This adds extra uncertainty!
- The statistic $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ does NOT follow $N(0,1)$
  - It follows the t-distribution with $n-1$ degrees of freedom
- Properties of t-distribution compared to Normal:
  - Also symmetric and bell-shaped
  - **Heavier tails** $\Rightarrow$ larger critical values
  - t critical values $>$ z critical values
  - Results in **wider** confidence intervals
- As $n \to \infty$: $t_{n-1} \to N(0,1)$

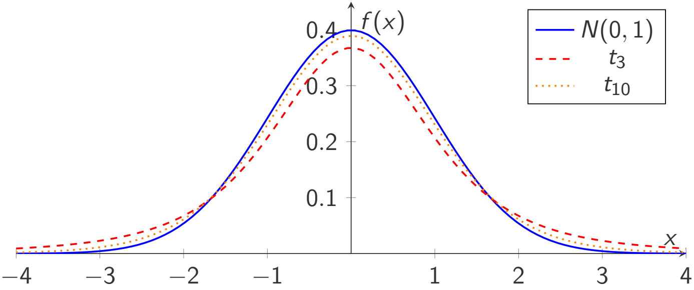
Comparing t and Normal Distributions



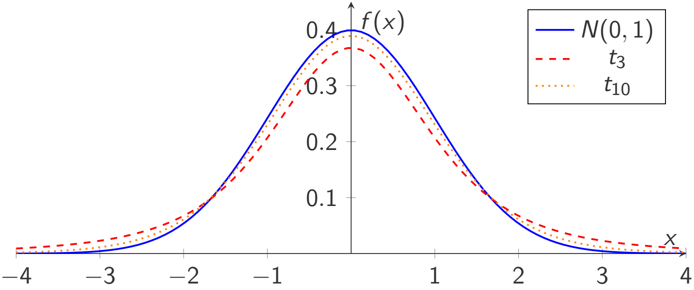
Comparing t and Normal Distributions



Comparing t and Normal Distributions



Comparing t and Normal Distributions



Key point: Lower df $\Rightarrow$ heavier tails $\Rightarrow$ larger critical values $\Rightarrow$ wider CI

CI Formulas: Mean (σ unknown) and Variance

CI for Mean μ (σ unknown) — the t-interval, Procedure 3.2:

$$\bar{x} - t_{\alpha/2,\,n-1}\frac{s}{\sqrt{n}} \;\leq\; \mu \;\leq\; \bar{x} + t_{\alpha/2,\,n-1}\frac{s}{\sqrt{n}} \tag{3.5}$$

CI Formulas: Mean (σ unknown) and Variance

CI for Mean μ (σ unknown) — the t-interval, Procedure 3.2:

$$\bar{x} - t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}} \;\le\; \mu \;\le\; \bar{x} + t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}} \tag{3.5}$$

CI for Variance σ^2 — the χ^2-interval, Procedure 3.3:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,\, n-1}} \;\le\; \sigma^2 \;\le\; \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\, n-1}} \tag{3.6}$$

CI Formulas: Mean (σ unknown) and Variance

CI for Mean μ (σ unknown) — the t-interval, Procedure 3.2:

$$\bar{x} - t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}} \;\le\; \mu \;\le\; \bar{x} + t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}} \tag{3.5}$$

CI for Variance σ^2 — the χ^2-interval, Procedure 3.3:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,\, n-1}} \;\le\; \sigma^2 \;\le\; \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\, n-1}} \tag{3.6}$$

For a CI on σ, take square roots of both endpoints of (3.6).

CI Formulas: Proportion

CI for Proportion p (large sample) — Procedure 3.4:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;\leq\; p \;\leq\; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{3.7}$$

CI Formulas: Proportion

CI for Proportion p (large sample) — Procedure 3.4:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;\leq\; p \;\leq\; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{3.7}$$

Sample Size for Proportion:

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p}) \tag{3.8}$$

CI Formulas: Proportion

CI for Proportion p (large sample) — Procedure 3.4:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;\leq\; p \;\leq\; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{3.7}$$

Sample Size for Proportion:

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p}) \tag{3.8}$$

If no prior $\hat{p}$ is available, use $\hat{p} = 0.5$ — the conservative version, (3.9) in the textbook.

t-Distribution Critical Values, Table 3.3 ($t_{\alpha,\nu}$ = value where $P(T > t_{\alpha,\nu}) = \alpha$)

df (ν)	Upper-tail probability α			
	0.10	**0.05**	**0.025**	**0.01**
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
5	1.476	2.015	2.571	3.365
10	1.372	1.812	2.228	2.764
15	1.341	1.753	2.131	2.602
20	1.325	1.725	2.086	2.528
30	1.310	1.697	2.042	2.457
∞	1.282	1.645	1.960	2.326

t-Distribution Critical Values, Table 3.3 ($t_{\alpha,\nu}$ = value where $P(T > t_{\alpha,\nu}) = \alpha$)

	Upper-tail probability α			
df (ν)	**0.10**	**0.05**	**0.025**	**0.01**
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
5	1.476	2.015	2.571	3.365
10	1.372	1.812	2.228	2.764
15	1.341	1.753	2.131	2.602
20	1.325	1.725	2.086	2.528
30	1.310	1.697	2.042	2.457
∞	1.282	1.645	1.960	2.326

Example: For 95% two-sided CI with $n = 11$ samples (df = 10): Use $t_{0.025,10} = 2.228$

χ^2-Distribution Critical Values, Table 3.4 ($\chi^2_{\alpha,\nu}$ = value s.t. $P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$)

	Upper-tail probability α			
df (ν)	**0.975**	**0.95**	**0.05**	**0.025**
1	0.001	0.004	3.841	5.024
2	0.051	0.103	5.991	7.378
5	0.831	1.145	11.070	12.833
9	2.700	3.325	16.919	19.023
10	3.247	3.940	18.307	20.483
19	8.907	10.117	30.144	32.852
20	9.591	10.851	31.410	34.170

χ^2-Distribution Critical Values, Table 3.4 ($\chi^2_{\alpha,\nu}$ = value s.t. $P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$)

	Upper-tail probability α			
df (ν)	**0.975**	**0.95**	**0.05**	**0.025**
1	0.001	0.004	3.841	5.024
2	0.051	0.103	5.991	7.378
5	0.831	1.145	11.070	12.833
9	2.700	3.325	16.919	19.023
10	3.247	3.940	18.307	20.483
19	8.907	10.117	30.144	32.852
20	9.591	10.851	31.410	34.170

Example: For 95% CI for σ^2 with $n = 10$ ($df = 9$): Use $\chi^2_{0.975,9} = 2.700$ and $\chi^2_{0.025,9} = 19.023$

Summary

- When σ is **unknown**, use the *t*-**distribution** (3.5):
  - Replace σ with s
  - Replace $z_{\alpha/2}$ with $t_{\alpha/2,n-1}$
  - Results in wider CI (accounts for extra uncertainty)

Summary

- When σ is **unknown**, use the t-**distribution** (3.5):
  - Replace σ with s
  - Replace $z_{\alpha/2}$ with $t_{\alpha/2,n-1}$
  - Results in wider CI (accounts for extra uncertainty)
- For **variance** σ^2, use the χ^2-**distribution** (3.6):
  - Not symmetric—need two critical values
  - Very sensitive to Normality assumption

Summary

- When σ is **unknown**, use the t**-distribution** (3.5):
  - Replace σ with s
  - Replace $z_{\alpha/2}$ with $t_{\alpha/2,n-1}$
  - Results in wider CI (accounts for extra uncertainty)
- For **variance** σ^2, use the χ^2**-distribution** (3.6):
  - Not symmetric—need two critical values
  - Very sensitive to Normality assumption
- For **proportion** p, use **Normal approximation** (3.7):
  - Requires large sample: $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$
  - Standard error: $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Questions you might be asked about confidence intervals

- Given sample data with s (not σ), construct a CI for μ

Questions you might be asked about confidence intervals

- Given sample data with s (not σ), construct a CI for μ
- Given sample variance s^2, construct a CI for σ^2 or σ

Questions you might be asked about confidence intervals

- Given sample data with s (not σ), construct a CI for μ
- Given sample variance s^2, construct a CI for σ^2 or σ
- Given number of successes and sample size, construct a CI for proportion p

Questions you might be asked about confidence intervals

- Given sample data with s (not σ), construct a CI for μ
- Given sample variance s^2, construct a CI for σ^2 or σ
- Given number of successes and sample size, construct a CI for proportion p
- Determine sample size needed for estimating a proportion with given precision

Questions you might be asked about confidence intervals

- Given sample data with s (not σ), construct a CI for μ
- Given sample variance s^2, construct a CI for σ^2 or σ
- Given number of successes and sample size, construct a CI for proportion p
- Determine sample size needed for estimating a proportion with given precision
- Choose the appropriate distribution (z, t, or χ^2) for a given problem

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)
- For **mean** μ:

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)
- For **mean** μ:
  - σ known $\Rightarrow$ use z-interval, Procedure 3.1

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)
- For **mean** μ:
  - σ known $\Rightarrow$ use z-interval, Procedure 3.1
  - σ unknown $\Rightarrow$ use t-interval with $\text{df} = n - 1$, Procedure 3.2

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)
- For **mean** μ:
  - σ known $\Rightarrow$ use z-interval, Procedure 3.1
  - σ unknown $\Rightarrow$ use t-interval with df $= n - 1$, Procedure 3.2
- For **variance** σ^2:

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)
- For **mean** μ:
  - σ known $\Rightarrow$ use z-interval, Procedure 3.1
  - σ unknown $\Rightarrow$ use t-interval with df $= n - 1$, Procedure 3.2
- For **variance** σ^2:
  - Use χ^2-interval, Procedure 3.3 (need both upper and lower critical values)

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)
- For **mean** μ:
  - σ known $\Rightarrow$ use z-interval, Procedure 3.1
  - σ unknown $\Rightarrow$ use t-interval with df $= n - 1$, Procedure 3.2
- For **variance** σ^2:
  - Use χ^2-interval, Procedure 3.3 (need both upper and lower critical values)
- For **proportion** p:

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)
- For **mean** μ:
  - σ known $\Rightarrow$ use z-interval, Procedure 3.1
  - σ unknown $\Rightarrow$ use t-interval with df $= n - 1$, Procedure 3.2
- For **variance** σ^2:
  - Use χ^2-interval, Procedure 3.3 (need both upper and lower critical values)
- For **proportion** p:
  - Check: $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$?

General Recipe: Procedure 3.5 (choosing the right interval)

- First identify: What parameter? (μ, σ^2, or p?)
- For **mean** μ:
  - σ known $\Rightarrow$ use z-interval, Procedure 3.1
  - σ unknown $\Rightarrow$ use t-interval with df $= n - 1$, Procedure 3.2
- For **variance** σ^2:
  - Use χ^2-interval, Procedure 3.3 (need both upper and lower critical values)
- For **proportion** p:
  - Check: $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$?
  - Use z-interval with SE $= \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, Procedure 3.4

Questions you might be asked about confidence intervals

What variants we might see?

- Given only summary statistics ($\bar{x}$, s, n) vs raw data

Questions you might be asked about confidence intervals

What variants we might see?

- Given only summary statistics ($\bar{x}$, s, n) vs raw data
  - Same formulas apply; may need to compute $\bar{x}$ and s first

Questions you might be asked about confidence intervals

What variants we might see?

- Given only summary statistics ($\bar{x}$, s, n) vs raw data
  - Same formulas apply; may need to compute $\bar{x}$ and s first
- One-sided bounds for t and proportion intervals

Questions you might be asked about confidence intervals

What variants we might see?

- Given only summary statistics ($\bar{x}$, s, n) vs raw data
  - Same formulas apply; may need to compute $\bar{x}$ and s first
- One-sided bounds for t and proportion intervals
  - Use t_α or z_α instead of $t_{\alpha/2}$ or $z_{\alpha/2}$

Questions you might be asked about confidence intervals

What variants we might see?

- Given only summary statistics ($\bar{x}$, s, n) vs raw data
    - Same formulas apply; may need to compute $\bar{x}$ and s first
- One-sided bounds for t and proportion intervals
    - Use t_α or z_α instead of $t_{\alpha/2}$ or $z_{\alpha/2}$
- CI for standard deviation σ (not σ^2)

Questions you might be asked about confidence intervals

What variants we might see?

- Given only summary statistics ($\bar{x}$, s, n) vs raw data
  - Same formulas apply; may need to compute $\bar{x}$ and s first
- One-sided bounds for t and proportion intervals
  - Use t_α or z_α instead of $t_{\alpha/2}$ or $z_{\alpha/2}$
- CI for standard deviation σ (not σ^2)
  - Take square root of the variance CI bounds

Questions you might be asked about confidence intervals

What variants we might see?

- Given only summary statistics ($\bar{x}$, s, n) vs raw data
  - Same formulas apply; may need to compute $\bar{x}$ and s first
- One-sided bounds for t and proportion intervals
  - Use t_α or z_α instead of $t_{\alpha/2}$ or $z_{\alpha/2}$
- CI for standard deviation σ (not σ^2)
  - Take square root of the variance CI bounds
- Sample size for proportion with no prior estimate

Questions you might be asked about confidence intervals

What variants we might see?

- Given only summary statistics ($\bar{x}$, s, n) vs raw data
  - Same formulas apply; may need to compute $\bar{x}$ and s first
- One-sided bounds for t and proportion intervals
  - Use t_α or z_α instead of $t_{\alpha/2}$ or $z_{\alpha/2}$
- CI for standard deviation σ (not σ^2)
  - Take square root of the variance CI bounds
- Sample size for proportion with no prior estimate
  - Use $\hat{p} = 0.5$ (conservative, maximizes $\hat{p}(1 - \hat{p})$, (3.9))

What confidence intervals we use in this class? (Table 3.7)

- Mean μ with σ known (z-interval, (3.1))

What confidence intervals we use in this class? (Table 3.7)

- Mean μ with σ known (z-interval, (3.1))
- Mean μ with σ unknown (t-interval, (3.5))

What confidence intervals we use in this class? (Table 3.7)

- Mean μ with σ known (z-interval, (3.1))
- Mean μ with σ unknown (t-interval, (3.5))
- Variance σ^2 (χ^2-interval, (3.6))

What confidence intervals we use in this class? (Table 3.7)

- Mean μ with σ known (z-interval, (3.1))
- Mean μ with σ unknown (t-interval, (3.5))
- Variance σ^2 (χ^2-interval, (3.6))
- Proportion p (z-interval with $\sqrt{\hat{p}(1-\hat{p})/n}$, (3.7))

What confidence intervals we use in this class? (Table 3.7)

- Mean μ with σ known (z-interval, (3.1))
- Mean μ with σ unknown (t-interval, (3.5))
- Variance σ^2 (χ^2-interval, (3.6))
- Proportion p (z-interval with $\sqrt{\hat{p}(1-\hat{p})/n}$, (3.7))
- (Later chapters) Difference in means $\mu_1 - \mu_2$

What confidence intervals we use in this class? (Table 3.7)

- Mean μ with σ known (z-interval, (3.1))
- Mean μ with σ unknown (t-interval, (3.5))
- Variance σ^2 (χ^2-interval, (3.6))
- Proportion p (z-interval with $\sqrt{\hat{p}(1-\hat{p})/n}$, (3.7))
- (Later chapters) Difference in means $\mu_1 - \mu_2$
- (Later chapters) Difference in proportions $p_1 - p_2$

Class Announcements

- Next class: Practice on confidence intervals

Class Announcements

- Next class: Practice on confidence intervals
- Check out recorded lectures on Blackboard if you need a refresher

Class Announcements

- Next class: Practice on confidence intervals
- Check out recorded lectures on Blackboard if you need a refresher
- Quiz #2 next Wednesday!

Class Announcements

- Next class: Practice on confidence intervals
- Check out recorded lectures on Blackboard if you need a refresher
- Quiz #2 next Wednesday!
- Homework #2 due next Wednesday (before class)

Class Announcements

- Next class: Practice on confidence intervals
- Check out recorded lectures on Blackboard if you need a refresher
- Quiz #2 next Wednesday!
- Homework #2 due next Wednesday (before class)
- Office hour: Tuesdays 9-10 AM (22-219 or Zoom)