ISE 315: Engineering Statistics

Lecture 8: Interval Toolkit Recap & Introduction to Hypothesis Testing

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 8

Interval Toolkit Recap & Introduction to Hypothesis Testing (Chapter 4)

Lecture 8 Outline

- Recap: the Chapter 3 interval toolkit
- Introduction to Hypothesis Testing (Chapter 4)
  - From confidence intervals to hypothesis tests
  - Statistical hypotheses: H_0 vs H_1
  - Type I and Type II errors
  - Test statistics and critical regions
- Quiz #2 (30 minutes)

Chapter 3 Recap

The Interval Toolkit: Four Confidence Intervals

The Four Intervals You Already Own

Chapter 3 in One Table

Parameter	**Condition**	**Two-Sided $100(1-\alpha)\%$ CI**	**Book Eq.**
μ	σ known	$\bar{x} \pm z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}$	(3.1)

Textbook: Chapter 3 ((3.1) onward); recap collected in Table 4.1

The Four Intervals You Already Own

Chapter 3 in One Table

Parameter	**Condition**	**Two-Sided** $100(1-\alpha)\%$ **CI**	**Book Eq.**
μ	σ known	$\bar{x} \pm z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}$	(3.1)
μ	σ unknown	$\bar{x} \pm t_{\alpha/2,\,n-1}\,\dfrac{s}{\sqrt{n}}$	(3.5)

Textbook: Chapter 3 ((3.1) onward); recap collected in Table 4.1

The Four Intervals You Already Own

Chapter 3 in One Table

Parameter	Condition	Two-Sided $100(1-\alpha)\%$ CI	Book Eq.
μ	σ known	$\bar{x} \pm z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}$	(3.1)
μ	σ unknown	$\bar{x} \pm t_{\alpha/2,\,n-1}\,\dfrac{s}{\sqrt{n}}$	(3.5)
σ^2	Normal pop.	$\left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}},\ \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}}\right)$	(3.6)

Textbook: Chapter 3 ((3.1) onward); recap collected in Table 4.1

The Four Intervals You Already Own

Chapter 3 in One Table

Parameter	**Condition**	**Two-Sided** $100(1-\alpha)\%$ **CI**	**Book Eq.**
μ	σ known	$\bar{x} \pm z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}$	(3.1)
μ	σ unknown	$\bar{x} \pm t_{\alpha/2,\,n-1}\,\dfrac{s}{\sqrt{n}}$	(3.5)
σ^2	Normal pop.	$\left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}},\ \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}}\right)$	(3.6)
p	Large n	$\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$	(3.7)

Textbook: Chapter 3 ((3.1) onward); recap collected in Table 4.1

The Four Intervals You Already Own

Chapter 3 in One Table

Parameter	Condition	Two-Sided $100(1-\alpha)\%$ CI	Book Eq.
μ	σ known	$\bar{x} \pm z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}$	(3.1)
μ	σ unknown	$\bar{x} \pm t_{\alpha/2,\,n-1}\,\dfrac{s}{\sqrt{n}}$	(3.5)
σ^2	Normal pop.	$\left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}},\ \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}}\right)$	(3.6)
p	Large n	$\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$	(3.7)

Today's claim: every one of these intervals has a matching hypothesis test.

Textbook: Chapter 3 ((3.1) onward); recap collected in Table 4.1

Which Interval When?

Three Questions (Procedure 3.5)

1. **Which parameter?** A mean μ, a variance σ^2, or a proportion p?

Which Interval When?

Three Questions (Procedure 3.5)

1. **Which parameter?** A mean μ, a variance σ^2, or a proportion p?
2. **For a mean: is σ known?**
   - Known $\Rightarrow$ z interval (3.1)
   - Unknown $\Rightarrow$ t interval (3.5) with $n - 1$ df

Which Interval When?

Three Questions (Procedure 3.5)

1. **Which parameter?** A mean μ, a variance σ^2, or a proportion p?
2. **For a mean: is σ known?**
   - Known $\Rightarrow$ z interval (3.1)
   - Unknown $\Rightarrow$ t interval (3.5) with $n-1$ df
3. **What is the data like?**
   - Variance needs a Normal population $\Rightarrow$ χ^2 interval (3.6)
   - Proportion needs a large sample $\Rightarrow$ z interval (3.7)

Which Interval When?

Three Questions (Procedure 3.5)

1. **Which parameter?** A mean μ, a variance σ^2, or a proportion p?
2. **For a mean: is σ known?**
   - Known $\Rightarrow$ z interval (3.1)
   - Unknown $\Rightarrow$ t interval (3.5) with $n-1$ df
3. **What is the data like?**
   - Variance needs a Normal population $\Rightarrow$ χ^2 interval (3.6)
   - Proportion needs a large sample $\Rightarrow$ z interval (3.7)

Keep these three questions. They pick the hypothesis test in Chapter 4 too.

Warm-Up: One Interval, Step by Step

AI Safety Testing Context

A red team runs $n = 25$ held-out safety-eval suites on a new LLM. The mean response latency is $\bar{x} = 310$ ms with $s = 40$ ms. Build a 95% CI for the true mean latency μ.

Warm-Up: One Interval, Step by Step

AI Safety Testing Context

A red team runs $n = 25$ held-out safety-eval suites on a new LLM. The mean response latency is $\bar{x} = 310$ ms with $s = 40$ ms. Build a 95% CI for the true mean latency μ.

Which interval? Mean μ, σ unknown $\Rightarrow$ t interval (3.5), df $= 25 - 1 = 24$

Warm-Up: One Interval, Step by Step

AI Safety Testing Context

A red team runs $n = 25$ held-out safety-eval suites on a new LLM. The mean response latency is $\bar{x} = 310$ ms with $s = 40$ ms. Build a 95% CI for the true mean latency μ.

Which interval? Mean μ, σ unknown $\Rightarrow$ t interval (3.5), df $= 25 - 1 = 24$

Table value: $t_{0.025,\,24} = 2.064$

Warm-Up: One Interval, Step by Step

AI Safety Testing Context

A red team runs $n = 25$ held-out safety-eval suites on a new LLM. The mean response latency is $\bar{x} = 310$ ms with $s = 40$ ms. Build a 95% CI for the true mean latency μ.

Which interval? Mean μ, σ unknown $\Rightarrow$ t interval (3.5), df $= 25 - 1 = 24$

Table value: $t_{0.025,\,24} = 2.064$

Substitute: $\bar{x} \pm t_{0.025,\,24}\,\dfrac{s}{\sqrt{n}} = 310 \pm 2.064 \times \dfrac{40}{\sqrt{25}} = 310 \pm 2.064 \times 8$

Warm-Up: One Interval, Step by Step

AI Safety Testing Context

A red team runs $n = 25$ held-out safety-eval suites on a new LLM. The mean response latency is $\bar{x} = 310$ ms with $s = 40$ ms. Build a 95% CI for the true mean latency μ.

Which interval? Mean μ, σ unknown $\Rightarrow$ t interval (3.5), df $= 25 - 1 = 24$

Table value: $t_{0.025,\,24} = 2.064$

Substitute: $\bar{x} \pm t_{0.025,\,24} \dfrac{s}{\sqrt{n}} = 310 \pm 2.064 \times \dfrac{40}{\sqrt{25}} = 310 \pm 2.064 \times 8$

Compute: 310 ± 16.51

Warm-Up: One Interval, Step by Step

AI Safety Testing Context

A red team runs $n = 25$ held-out safety-eval suites on a new LLM. The mean response latency is $\bar{x} = 310$ ms with $s = 40$ ms. Build a 95% CI for the true mean latency μ.

Which interval? Mean μ, σ unknown $\Rightarrow$ t interval (3.5), df $= 25 - 1 = 24$

Table value: $t_{0.025,\,24} = 2.064$

Substitute: $\bar{x} \pm t_{0.025,\,24} \dfrac{s}{\sqrt{n}} = 310 \pm 2.064 \times \dfrac{40}{\sqrt{25}} = 310 \pm 2.064 \times 8$

Compute: 310 ± 16.51

Interval: $(293.5,\ 326.5)$ ms. We're 95% confident the true mean latency is in this range.

Chapter 4

Tests of Hypotheses for a Single Sample

From Confidence Intervals to Hypothesis Testing

The Big Picture

Chapter 3
Confidence Intervals
"What is μ?"

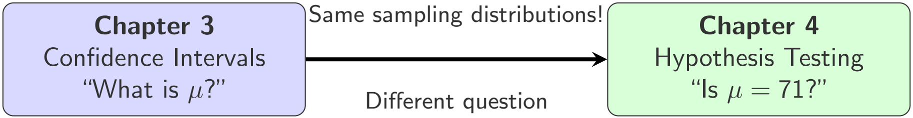
Same sampling distributions!
$\longrightarrow$
Different question

Chapter 4
Hypothesis Testing
"Is $\mu = 71$?"

Textbook: Chapter 4 ((4.1) onward)

From Confidence Intervals to Hypothesis Testing

The Big Picture



- **Chapter 3 (CI):** Estimate an unknown parameter
  - "The true mean container dwell time is between 72.08 and 79.92 hours"

Textbook: Chapter 4 ((4.1) onward)

From Confidence Intervals to Hypothesis Testing

The Big Picture

Chapter 3
Confidence Intervals
"What is μ?"

Same sampling distributions!

Different question

Chapter 4
Hypothesis Testing
"Is $\mu = 71$?"

- **Chapter 3 (CI):** Estimate an unknown parameter
  - "The true mean container dwell time is between 72.08 and 79.92 hours"
- **Chapter 4 (HT):** Test a claim about a parameter
  - "Is the true mean dwell time equal to 71 hours?"

Textbook: Chapter 4 ((4.1) onward)

Motivating Example: Container Dwell Time

Supply Chain Context

Scenario: A logistics analyst measures container dwell times at a port terminal.

- Sample: $n = 64$ containers, $\bar{x} = 76$ hours, known $\sigma = 16$ hours

Motivating Example: Container Dwell Time

Supply Chain Context

Scenario: A logistics analyst measures container dwell times at a port terminal.

- Sample: $n = 64$ containers, $\bar{x} = 76$ hours, known $\sigma = 16$ hours

Chapter 3 Question: What is the true mean dwell time?

- 95% CI from (3.1): $(72.08, \ 79.92)$ hours

Motivating Example: Container Dwell Time

Supply Chain Context

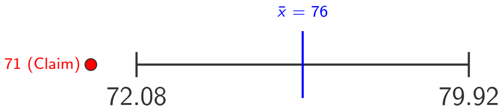
Scenario: A logistics analyst measures container dwell times at a port terminal.

- Sample: $n = 64$ containers, $\bar{x} = 76$ hours, known $\sigma = 16$ hours

Chapter 3 Question: What is the true mean dwell time?

- 95% CI from (3.1): $(72.08, 79.92)$ hours

Chapter 4 Question: The terminal operator claims the mean dwell time is 71 hours. Is this claim supported by our data?



Motivating Example: Container Dwell Time

Supply Chain Context

Scenario: A logistics analyst measures container dwell times at a port terminal.

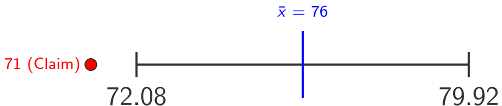
- Sample: $n = 64$ containers, $\bar{x} = 76$ hours, known $\sigma = 16$ hours

Chapter 3 Question: What is the true mean dwell time?

- 95% CI from (3.1): (72.08, 79.92) hours

Chapter 4 Question: The terminal operator claims the mean dwell time is 71 hours. Is this claim supported by our data?

- 71 is **outside** our 95% CI! (the claim might be false?)



What is Hypothesis Testing?

Definition: A procedure for making decisions about population parameters based on sample data.

What is Hypothesis Testing?

Definition: A procedure for making decisions about population parameters based on sample data.

The Process:

What is Hypothesis Testing?

Definition: A procedure for making decisions about population parameters based on sample data.

The Process:

1. State a **hypothesis** (claim) about the parameter

What is Hypothesis Testing?

Definition: A procedure for making decisions about population parameters based on sample data.

The Process:

1. State a **hypothesis** (claim) about the parameter
2. Collect sample data

What is Hypothesis Testing?

Definition: A procedure for making decisions about population parameters based on sample data.

The Process:

1. State a **hypothesis** (claim) about the parameter
2. Collect sample data
3. Calculate a **test statistic**

What is Hypothesis Testing?

Definition: A procedure for making decisions about population parameters based on sample data.

The Process:

1. State a **hypothesis** (claim) about the parameter
2. Collect sample data
3. Calculate a **test statistic**
4. Make a **decision**: Reject or fail to reject the hypothesis

What is Hypothesis Testing?

Definition: A procedure for making decisions about population parameters based on sample data.

The Process:

1. State a **hypothesis** (claim) about the parameter
2. Collect sample data
3. Calculate a **test statistic**
4. Make a **decision**: Reject or fail to reject the hypothesis

Key Idea: We use probability to quantify the evidence against the hypothesis.

The Two Hypotheses

H_0 and H_1

The standard two-sided pair for a mean:

$$H_0 : \ \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \ \mu \neq \mu_0 \tag{4.1}$$

The Two Hypotheses

H_0 and H_1

The standard two-sided pair for a mean:

$$H_0: \ \mu = \mu_0 \qquad \text{versus} \qquad H_1: \ \mu \neq \mu_0 \tag{4.1}$$

Null Hypothesis H_0:

- The "default" or "status quo"
- What we assume is true
- Usually contains "$=$"
- Example: $H_0 : \mu = 71$

The Two Hypotheses

H_0 and H_1

The standard two-sided pair for a mean:

$$H_0: \ \mu = \mu_0 \qquad \text{versus} \qquad H_1: \ \mu \neq \mu_0 \tag{4.1}$$

Null Hypothesis H_0:

- The "default" or "status quo"
- What we assume is true
- Usually contains "$=$"
- Example: $H_0 : \mu = 71$

Alternative Hypothesis H_1:

- The "research hypothesis"
- What we want to prove
- Contains "$\neq$", "$<$", or "$>$"
- Example: $H_1 : \mu \neq 71$

The Two Hypotheses

H_0 and H_1

The standard two-sided pair for a mean:

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu \neq \mu_0 \tag{4.1}$$

Null Hypothesis H_0:

- The “default” or “status quo”
- What we assume is true
- Usually contains “$=$”
- Example: $H_0 : \mu = 71$

Alternative Hypothesis H_1:

- The “research hypothesis”
- What we want to prove
- Contains “$\neq$”, “$<$”, or “$>$”
- Example: $H_1 : \mu \neq 71$

Important: We never “accept” H_0.

Three Types of Hypothesis Tests

One Pair (4.1), Three Alternatives

Type	**Hypotheses**	**Example**
Two-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$	Is mean dwell time different from 71 h?

Three Types of Hypothesis Tests

One Pair (4.1), Three Alternatives

Type	Hypotheses	Example
Two-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$	Is mean dwell time different from 71 h?
Left-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$	Is a startup's mean churn below target?

Three Types of Hypothesis Tests

One Pair (4.1), Three Alternatives

Type	**Hypotheses**	**Example**
Two-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$	Is mean dwell time different from 71 h?
Left-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$	Is a startup's mean churn below target?
Right-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$	Is an AV's mean disengagement rate above spec?

Three Types of Hypothesis Tests

One Pair (4.1), Three Alternatives

Type	Hypotheses	Example
Two-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$	Is mean dwell time different from 71 h?
Left-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$	Is a startup's mean churn below target?
Right-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$	Is an AV's mean disengagement rate above spec?

Can you guess the connection to one- or two-sided CIs?

Type I and Type II Errors

Two Ways to Be Wrong (Table 4.2)

Decision	**Truth (Unknown)**	
	H_0 is True	H_0 is False

Type I and Type II Errors

Two Ways to Be Wrong (Table 4.2)

	Truth (Unknown)	
Decision	H_0 is True	H_0 is False
Reject H_0	**Type I Error** (α)	Correct!

Type I and Type II Errors

Two Ways to Be Wrong (Table 4.2)

	Truth (Unknown)	
Decision	H_0 is True	H_0 is False
Reject H_0	**Type I Error** (α)	Correct!
Fail to Reject H_0	Correct!	**Type II Error** (β)

Type I and Type II Errors

Two Ways to Be Wrong (Table 4.2)

Decision	**Truth (Unknown)**	
	H_0 is True	H_0 is False
Reject H_0	**Type I Error** (α)	Correct!
Fail to Reject H_0	Correct!	**Type II Error** (β)

The two error probabilities have book names:

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true}), \qquad \beta = P(\text{fail to reject } H_0 \mid H_0 \text{ false}) \tag{4.2}$$

Type I and Type II Errors

Two Ways to Be Wrong (Table 4.2)

Decision	Truth (Unknown)	
	H_0 is True	H_0 is False
Reject H_0	**Type I Error** (α)	Correct!
Fail to Reject H_0	Correct!	**Type II Error** (β)

The two error probabilities have book names:

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true}), \qquad \beta = P(\text{fail to reject } H_0 \mid H_0 \text{ false}) \tag{4.2}$$

- **Type I (α):** "false positive" or "false alarm"

Type I and Type II Errors

Two Ways to Be Wrong (Table 4.2)

	Truth (Unknown)	
Decision	H_0 is True	H_0 is False
Reject H_0	**Type I Error** (α)	Correct!
Fail to Reject H_0	Correct!	**Type II Error** (β)

The two error probabilities have book names:

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true}), \qquad \beta = P(\text{fail to reject } H_0 \mid H_0 \text{ false}) \tag{4.2}$$

- **Type I (α):** "false positive" or "false alarm"
- **Type II (β):** "false negative" or "missed detection"

Analogy: The AI Safety Release Gate

Type I vs Type II Errors

Scenario: Deciding whether a perception model is safe enough to ship.

- H_0: Model is **safe** (failure rate meets the spec)
- H_1: Model is **unsafe** (failure rate exceeds the spec)

Analogy: The AI Safety Release Gate

Type I vs Type II Errors

Scenario: Deciding whether a perception model is safe enough to ship.

- H_0: Model is **safe** (failure rate meets the spec)
- H_1: Model is **unsafe** (failure rate exceeds the spec)

Type I Error
(Block a safe model)

- Retraining, delayed launch
- Cost: SAR 200,000 in rework

Analogy: The AI Safety Release Gate

Type I vs Type II Errors

Scenario: Deciding whether a perception model is safe enough to ship.

- H_0: Model is **safe** (failure rate meets the spec)
- H_1: Model is **unsafe** (failure rate exceeds the spec)

Type I Error
(Block a safe model)

- Retraining, delayed launch
- Cost: SAR 200,000 in rework

Type II Error
(Ship an unsafe model)

- Road incidents, recalls, audits
- Cost: SAR millions + trust

Analogy: The AI Safety Release Gate

Type I vs Type II Errors

Scenario: Deciding whether a perception model is safe enough to ship.

- H_0: Model is **safe** (failure rate meets the spec)
- H_1: Model is **unsafe** (failure rate exceeds the spec)

Type I Error
(Block a safe model)

- Retraining, delayed launch
- Cost: SAR 200,000 in rework

Type II Error
(Ship an unsafe model)

- Road incidents, recalls, audits
- Cost: SAR millions + trust

Trade-off: Reducing α typically increases β (and vice versa) — see (4.2).

Significance Level α

Controlling Type I Error

Definition: The significance level α is the maximum probability of Type I error we're willing to accept.

Significance Level α

Controlling Type I Error

Definition: The significance level α is the maximum probability of Type I error we're willing to accept.

Common choices:

- $\alpha = 0.10$ (10%) – Less stringent (early startup A/B tests)
- $\alpha = 0.05$ (5%) – Most common
- $\alpha = 0.01$ (1%) – More stringent (safety-critical releases)

Significance Level α

Controlling Type I Error

Definition: The significance level α is the maximum probability of Type I error we're willing to accept.

Common choices:

- $\alpha = 0.10$ (10%) – Less stringent (early startup A/B tests)
- $\alpha = 0.05$ (5%) – Most common
- $\alpha = 0.01$ (1%) – More stringent (safety-critical releases)

You have seen this in Chapter 3: $\alpha = 0.05$ for a hypothesis test $\leftrightarrow$ 95% CI

The Test Statistic

Measuring Evidence Against H_0

Idea: Convert sample data into a single number that measures how far the data is from what H_0 predicts.

The Test Statistic

Measuring Evidence Against H_0

Idea: Convert sample data into a single number that measures how far the data is from what H_0 predicts.

For testing μ when σ is known, the book gives the Z-test statistic:

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \tag{4.3}$$

- $\bar{X}$ = sample mean (what we observed)
- μ_0 = hypothesized value (from H_0)
- $\sigma/\sqrt{n}$ = standard error

The Test Statistic

Measuring Evidence Against H_0

Idea: Convert sample data into a single number that measures how far the data is from what H_0 predicts.

For testing μ when σ is known, the book gives the Z-test statistic:

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \tag{4.3}$$

- $\bar{X}$ = sample mean (what we observed)
- μ_0 = hypothesized value (from H_0)
- $\sigma/\sqrt{n}$ = standard error

Interpretation: Z_0 tells us how many standard errors $\bar{X}$ is from μ_0.

Example: Testing the Operator's Claim

Container Dwell Time

Given: $n = 64$, $\bar{x} = 76$ hours, $\sigma = 16$ hours
Operator claims: $\mu = 71$ hours

Example: Testing the Operator's Claim

Container Dwell Time

Given: $n = 64$, $\bar{x} = 76$ hours, $\sigma = 16$ hours
Operator claims: $\mu = 71$ hours

Hypotheses (4.1): $H_0 : \mu = 71$ vs $H_1 : \mu \neq 71$

Example: Testing the Operator's Claim

Container Dwell Time

Given: $n = 64$, $\bar{x} = 76$ hours, $\sigma = 16$ hours
Operator claims: $\mu = 71$ hours

Hypotheses (4.1): $H_0 : \mu = 71$ vs $H_1 : \mu \neq 71$

Test statistic (4.3): $Z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{76 - 71}{16/\sqrt{64}} = \dfrac{5}{2} = 2.5$

Example: Testing the Operator's Claim

Container Dwell Time

Given: $n = 64$, $\bar{x} = 76$ hours, $\sigma = 16$ hours
Operator claims: $\mu = 71$ hours

Hypotheses (4.1): $H_0 : \mu = 71$ vs $H_1 : \mu \neq 71$

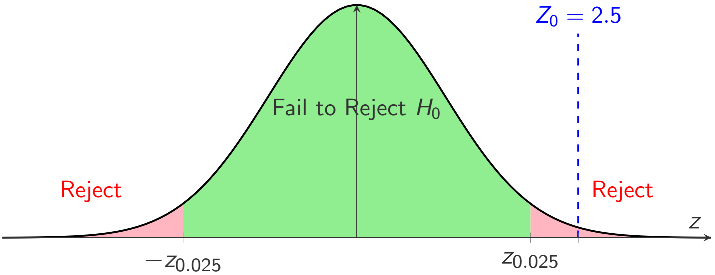
Test statistic (4.3): $Z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{76 - 71}{16/\sqrt{64}} = \dfrac{5}{2} = 2.5$

Question: Is $Z_0 = 2.5$ large enough to reject H_0?

We need a **decision rule**... (next slide!)

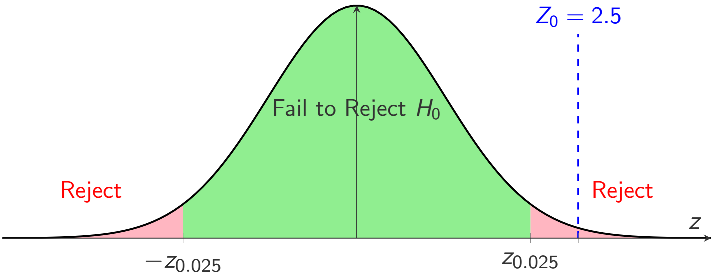
Critical Region and Decision Rule

See also Figure 4.2 in the textbook



Critical Region and Decision Rule

See also Figure 4.2 in the textbook

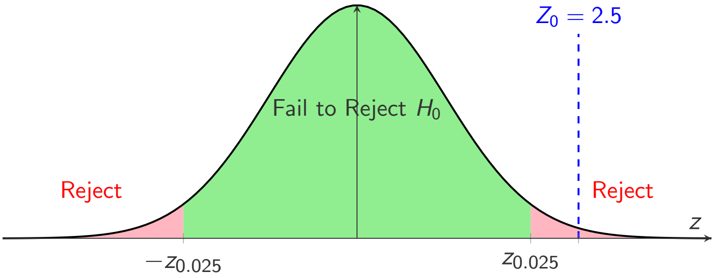


Decision Rule (for $\alpha = 0.05$, two-tailed): Reject H_0 if $|Z_0| > z_{0.025} = 1.96$

- Since $|2.5| > 1.96$, we **reject** H_0

Critical Region and Decision Rule

See also Figure 4.2 in the textbook



Decision Rule (for $\alpha = 0.05$, two-tailed): Reject H_0 if $|Z_0| > z_{0.025} = 1.96$

- Since $|2.5| > 1.96$, we **reject** H_0

Conclusion: The data provides sufficient evidence that the operator's claim is false.

Chapter 4: Key Takeaways So Far

- **Hypothesis testing** is about making decisions based on sample data

Chapter 4: Key Takeaways So Far

- **Hypothesis testing** is about making decisions based on sample data
- H_0 **(null):** The default claim we test against — pair (4.1)

Chapter 4: Key Takeaways So Far

- **Hypothesis testing** is about making decisions based on sample data
- H_0 **(null):** The default claim we test against — pair (4.1)
- H_1 **(alternative):** What we're trying to prove

Chapter 4: Key Takeaways So Far

- **Hypothesis testing** is about making decisions based on sample data
- H_0 **(null):** The default claim we test against — pair (4.1)
- H_1 **(alternative):** What we're trying to prove
- **Type I error (**α**):** Rejecting a true H_0 (false positive) — (4.2)

Chapter 4: Key Takeaways So Far

- **Hypothesis testing** is about making decisions based on sample data
- H_0 **(null):** The default claim we test against — pair (4.1)
- H_1 **(alternative):** What we're trying to prove
- **Type I error (α):** Rejecting a true H_0 (false positive) — (4.2)
- **Type II error (β):** Failing to reject a false H_0 (false negative)

Chapter 4: Key Takeaways So Far

- **Hypothesis testing** is about making decisions based on sample data
- H_0 **(null):** The default claim we test against — pair (4.1)
- H_1 **(alternative):** What we're trying to prove
- **Type I error (α):** Rejecting a true H_0 (false positive) — (4.2)
- **Type II error (β):** Failing to reject a false H_0 (false negative)
- **Test statistic:** Measures evidence against H_0 — e.g. Z_0 in (4.3)

Chapter 4: Key Takeaways So Far

- **Hypothesis testing** is about making decisions based on sample data
- H_0 **(null):** The default claim we test against — pair (4.1)
- H_1 **(alternative):** What we're trying to prove
- **Type I error (α):** Rejecting a true H_0 (false positive) — (4.2)
- **Type II error (β):** Failing to reject a false H_0 (false negative)
- **Test statistic:** Measures evidence against H_0 — e.g. Z_0 in (4.3)
- **Decision:** Compare test statistic to critical value

Chapter 4: Key Takeaways So Far

- **Hypothesis testing** is about making decisions based on sample data
- H_0 **(null):** The default claim we test against — pair (4.1)
- H_1 **(alternative):** What we're trying to prove
- **Type I error (α):** Rejecting a true H_0 (false positive) — (4.2)
- **Type II error (β):** Failing to reject a false H_0 (false negative)
- **Test statistic:** Measures evidence against H_0 — e.g. Z_0 in (4.3)
- **Decision:** Compare test statistic to critical value

Next lecture: More examples, P-values, and tests for σ unknown

Coming Up in Chapter 4...

Parameter	Test Statistic	Book Eq.
Mean μ (σ known)	$Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	(4.3)
Mean μ (σ unknown)	$T_0 = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$	(4.6)
Variance σ^2	$\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	(4.7)
Proportion p	$Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	(4.8)

Coming Up in Chapter 4...

Parameter	Test Statistic	Book Eq.
Mean μ (σ known)	$Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	(4.3)
Mean μ (σ unknown)	$T_0 = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$	(4.6)
Variance σ^2	$\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	(4.7)
Proportion p	$Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	(4.8)

Notice: Same distributions as the Chapter 3 confidence intervals!