ISE 315: Engineering Statistics

Lecture 9: Hypothesis Testing for a Single Sample

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 9

Hypothesis Testing for a Single Sample (Chapter 4, cont'd)

Lecture 9 Outline

- Announcements:
  - HW #3 due next week
  - Major Exam 1 in Week 8 (Chapters 2–5)
- Review: Key concepts from Lecture 8
- P-values in Hypothesis Tests
- The Seven-Step Testing Procedure
- Tests on the Mean: σ known and σ unknown

Quick Review

What We Learned in Lecture 8

Lecture 8 Recap: The Big Ideas

- **Hypothesis testing**: Use sample data to decide if a claim about a parameter is supported

Lecture 8 Recap: The Big Ideas

- **Hypothesis testing**: Use sample data to decide if a claim about a parameter is supported
- H_0 **(null)**: The default claim, usually with "$=$" — pair (4.1)

Lecture 8 Recap: The Big Ideas

- **Hypothesis testing**: Use sample data to decide if a claim about a parameter is supported
- H_0 **(null)**: The default claim, usually with "$=$" — pair (4.1)
- H_1 **(alternative)**: What we want to show, with "$\neq$", "$<$", or "$>$"

Lecture 8 Recap: The Big Ideas

- **Hypothesis testing**: Use sample data to decide if a claim about a parameter is supported
- H_0 **(null)**: The default claim, usually with "$=$" — pair (4.1)
- H_1 **(alternative)**: What we want to show, with "$\neq$", "$<$", or "$>$"
- **Test statistic**: $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ from (4.3) measures evidence against H_0

Lecture 8 Recap: The Big Ideas

- **Hypothesis testing**: Use sample data to decide if a claim about a parameter is supported
- H_0 **(null)**: The default claim, usually with "$=$" — pair (4.1)
- H_1 **(alternative)**: What we want to show, with "$\neq$", "$<$", or "$>$"
- **Test statistic**: $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ from (4.3) measures evidence against H_0
- **Type I error (α)**: Reject H_0 when true (false alarm) — (4.2)

Lecture 8 Recap: The Big Ideas

- **Hypothesis testing**: Use sample data to decide if a claim about a parameter is supported
- H_0 **(null)**: The default claim, usually with "$=$" — pair (4.1)
- H_1 **(alternative)**: What we want to show, with "$\neq$", "$<$", or "$>$"
- **Test statistic**: $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ from (4.3) measures evidence against H_0
- **Type I error (α)**: Reject H_0 when true (false alarm) — (4.2)
- **Type II error (β)**: Fail to reject H_0 when false (missed detection)

Lecture 8 Recap: The Big Ideas

- **Hypothesis testing**: Use sample data to decide if a claim about a parameter is supported
- H_0 **(null)**: The default claim, usually with "$=$" — pair (4.1)
- H_1 **(alternative)**: What we want to show, with "$\neq$", "$<$", or "$>$"
- **Test statistic**: $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ from (4.3) measures evidence against H_0
- **Type I error (α)**: Reject H_0 when true (false alarm) — (4.2)
- **Type II error (β)**: Fail to reject H_0 when false (missed detection)
- **Decision rule**: Reject H_0 if the test statistic falls in the **critical region**

Lecture 8 Recap: The Big Ideas

- **Hypothesis testing**: Use sample data to decide if a claim about a parameter is supported
- H_0 **(null)**: The default claim, usually with "$=$" — pair (4.1)
- H_1 **(alternative)**: What we want to show, with "$\neq$", "$<$", or "$>$"
- **Test statistic**: $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ from (4.3) measures evidence against H_0
- **Type I error (α)**: Reject H_0 when true (false alarm) — (4.2)
- **Type II error (β)**: Fail to reject H_0 when false (missed detection)
- **Decision rule**: Reject H_0 if the test statistic falls in the **critical region**

Today: We add P-values, the seven-step procedure, and tests when σ is unknown.

Section 4.5

P-Values in Hypothesis Tests

Two Approaches to Making a Decision

Approach 1: Critical Value (Fixed Significance Level)

- Choose α beforehand (e.g., $\alpha = 0.05$)
- Find critical value(s) from the distribution table
- Compare test statistic to critical value
- Decision: Reject or fail to reject

Textbook: Chapter 4, §4.5 ((4.4) onward)

Two Approaches to Making a Decision

Approach 1: Critical Value (Fixed Significance Level)

- Choose α beforehand (e.g., $\alpha = 0.05$)
- Find critical value(s) from the distribution table
- Compare test statistic to critical value
- Decision: Reject or fail to reject

Approach 2: P-Value (Procedure 4.1)

- Compute the probability of observing a test statistic *at least as extreme* as the one calculated, assuming H_0 is true
- Compare P-value to α
- Both approaches give the **same conclusion**!

Textbook: Chapter 4, §4.5 ((4.4) onward)

What is a P-Value?

Definition: The P-value is the *smallest* significance level at which we would reject H_0 given the observed data. The book writes it as

$$\text{P-value} = P\Big(\text{test statistic at least as extreme as observed} \,\Big|\, H_0 \text{ true}\Big) \tag{4.4}$$

What is a P-Value?

Definition: The P-value is the *smallest* significance level at which we would reject H_0 given the observed data. The book writes it as

$$\text{P-value} = P\big(\text{test statistic at least as extreme as observed} \,\big|\, H_0 \text{ true}\big) \qquad (4.4)$$

Decision Rule with the P-value (Procedure 4.1):

$$\boxed{\begin{array}{c} \text{If P-value} \leq \alpha \Rightarrow \textbf{Reject } H_0 \\[1em] \text{If P-value} > \alpha \Rightarrow \textbf{Fail to reject } H_0 \end{array}}$$

What is a P-Value?

Definition: The P-value is the *smallest* significance level at which we would reject H_0 given the observed data. The book writes it as

$$\text{P-value} = P\big(\text{test statistic at least as extreme as observed} \,\big|\, H_0 \text{ true}\big) \tag{4.4}$$

Decision Rule with the P-value (Procedure 4.1):

$$\boxed{\begin{array}{c} \text{If P-value} \le \alpha \Rightarrow \textbf{Reject } H_0 \\[1em] \text{If P-value} > \alpha \Rightarrow \textbf{Fail to reject } H_0 \end{array}}$$

Intuition: A small P-value means the data is very unlikely under H_0, so reject H_0.

Computing P-Values for the Z-Test

For the Z-test statistic Z_0 of (4.3), the book collects the three cases:

$$\text{P-value} = \begin{cases} 2\big(1 - \Phi(|Z_0|)\big) & H_1 : \mu \neq \mu_0 \\ 1 - \Phi(Z_0) & H_1 : \mu > \mu_0 \\ \Phi(Z_0) & H_1 : \mu < \mu_0 \end{cases} \tag{4.5}$$

Test Type	**Hypotheses**	**P-value**

Computing P-Values for the Z-Test

For the Z-test statistic Z_0 of (4.3), the book collects the three cases:

$$\text{P-value} = \begin{cases} 2\big(1 - \Phi(|Z_0|)\big) & H_1 : \mu \neq \mu_0 \\ 1 - \Phi(Z_0) & H_1 : \mu > \mu_0 \\ \Phi(Z_0) & H_1 : \mu < \mu_0 \end{cases} \tag{4.5}$$

Test Type	**Hypotheses**	**P-value**
Two-tailed	$H_1 : \mu \neq \mu_0$	$2\,P(Z > \lvert Z_0\rvert)$

Computing P-Values for the Z-Test

For the Z-test statistic Z_0 of (4.3), the book collects the three cases:

$$\text{P-value} = \begin{cases} 2\big(1 - \Phi(|Z_0|)\big) & H_1 : \mu \neq \mu_0 \\ 1 - \Phi(Z_0) & H_1 : \mu > \mu_0 \\ \Phi(Z_0) & H_1 : \mu < \mu_0 \end{cases} \tag{4.5}$$

Test Type	**Hypotheses**	**P-value**
Two-tailed	$H_1 : \mu \neq \mu_0$	$2\,P(Z > \|Z_0\|)$
Right-tailed	$H_1 : \mu > \mu_0$	$P(Z > Z_0)$

Computing P-Values for the Z-Test

For the Z-test statistic Z_0 of (4.3), the book collects the three cases:

$$\text{P-value} = \begin{cases} 2\big(1 - \Phi(|Z_0|)\big) & H_1 : \mu \neq \mu_0 \\ 1 - \Phi(Z_0) & H_1 : \mu > \mu_0 \\ \Phi(Z_0) & H_1 : \mu < \mu_0 \end{cases} \tag{4.5}$$

Test Type	**Hypotheses**	**P-value**
Two-tailed	$H_1 : \mu \neq \mu_0$	$2\,P(Z > \lvert Z_0 \rvert)$
Right-tailed	$H_1 : \mu > \mu_0$	$P(Z > Z_0)$
Left-tailed	$H_1 : \mu < \mu_0$	$P(Z < Z_0)$

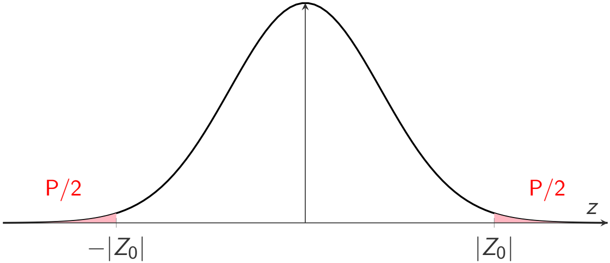
Computing P-Values for the Z-Test

For the Z-test statistic Z_0 of (4.3), the book collects the three cases:

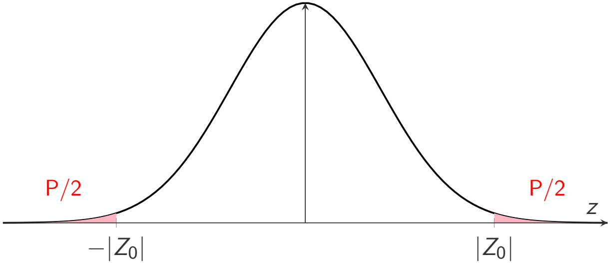
$$\text{P-value} = \begin{cases} 2\bigl(1 - \Phi(|Z_0|)\bigr) & H_1 : \mu \neq \mu_0 \\ 1 - \Phi(Z_0) & H_1 : \mu > \mu_0 \\ \Phi(Z_0) & H_1 : \mu < \mu_0 \end{cases} \tag{4.5}$$

Test Type	**Hypotheses**	**P-value**
Two-tailed	$H_1 : \mu \neq \mu_0$	$2\,P(Z > \|Z_0\|)$
Right-tailed	$H_1 : \mu > \mu_0$	$P(Z > Z_0)$
Left-tailed	$H_1 : \mu < \mu_0$	$P(Z < Z_0)$

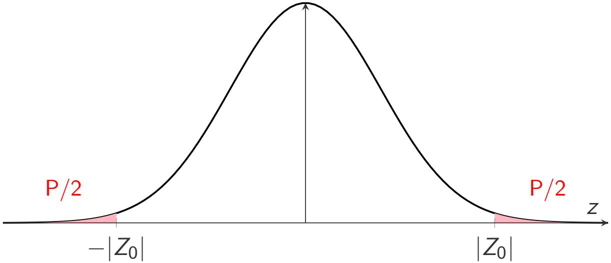
P-Value Illustrated: Two-Tailed Test



P-Value Illustrated: Two-Tailed Test

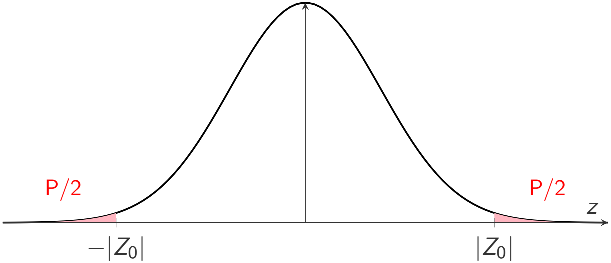


P-Value Illustrated: Two-Tailed Test



P-value $=$ total shaded area $= 2 \times P(Z > |Z_0|)$, the first case of (4.5)

P-Value Illustrated: Two-Tailed Test



P-value $=$ total shaded area $= 2 \times P(Z > |Z_0|)$, the first case of (4.5)

$=$ probability of observing data as or more extreme, **if** H_0 **is true**.

Revisiting the Operator's Claim with a P-Value

Example: Container Dwell Time

Recall: $n = 64$, $\bar{x} = 76$ h, $\sigma = 16$ h, $H_0 : \mu = 71$, $H_1 : \mu \neq 71$

Revisiting the Operator's Claim with a P-Value

Example: Container Dwell Time

Recall: $n = 64$, $\bar{x} = 76$ h, $\sigma = 16$ h, $H_0 : \mu = 71$, $H_1 : \mu \neq 71$

Test statistic: $Z_0 = 2.5$ (computed in Lecture 8 via (4.3))

Revisiting the Operator's Claim with a P-Value

Example: Container Dwell Time

Recall: $n = 64$, $\bar{x} = 76$ h, $\sigma = 16$ h, $H_0 : \mu = 71$, $H_1 : \mu \neq 71$

Test statistic: $Z_0 = 2.5$ (computed in Lecture 8 via (4.3))

P-value (two-tailed case of (4.5)):

$$\text{P-value} = 2\,P(Z > |2.5|) = 2\,(1 - \Phi(2.5))$$

Revisiting the Operator's Claim with a P-Value

Example: Container Dwell Time

Recall: $n = 64$, $\bar{x} = 76$ h, $\sigma = 16$ h, $H_0 : \mu = 71$, $H_1 : \mu \neq 71$

Test statistic: $Z_0 = 2.5$ (computed in Lecture 8 via (4.3))

P-value (two-tailed case of (4.5)):

$$\begin{aligned} \text{P-value} &= 2\,P(Z > |2.5|) = 2\,(1 - \Phi(2.5)) \\ &= 2\,(1 - 0.9938) \end{aligned}$$

Revisiting the Operator's Claim with a P-Value

Example: Container Dwell Time

Recall: $n = 64$, $\bar{x} = 76$ h, $\sigma = 16$ h, $H_0 : \mu = 71$, $H_1 : \mu \neq 71$

Test statistic: $Z_0 = 2.5$ (computed in Lecture 8 via (4.3))

P-value (two-tailed case of (4.5)):

$$\begin{aligned} \text{P-value} &= 2\,P(Z > |2.5|) = 2\,(1 - \Phi(2.5)) \\ &= 2\,(1 - 0.9938) \\ &= 2 \times 0.0062 = \boxed{0.0124} \end{aligned}$$

Revisiting the Operator's Claim with a P-Value

Example: Container Dwell Time

Recall: $n = 64$, $\bar{x} = 76$ h, $\sigma = 16$ h, $H_0 : \mu = 71$, $H_1 : \mu \neq 71$

Test statistic: $Z_0 = 2.5$ (computed in Lecture 8 via (4.3))

P-value (two-tailed case of (4.5)):

$$\begin{aligned} \text{P-value} &= 2\,P(Z > |2.5|) = 2\,(1 - \Phi(2.5)) \\ &= 2\,(1 - 0.9938) \\ &= 2 \times 0.0062 = \boxed{0.0124} \end{aligned}$$

Decision for $\alpha = 0.05$: P-value $= 0.0124 < 0.05 = \alpha$

Revisiting the Operator's Claim with a P-Value

Example: Container Dwell Time

Recall: $n = 64$, $\bar{x} = 76$ h, $\sigma = 16$ h, $H_0 : \mu = 71$, $H_1 : \mu \neq 71$

Test statistic: $Z_0 = 2.5$ (computed in Lecture 8 via (4.3))

P-value (two-tailed case of (4.5)):

$$\begin{aligned}\text{P-value} &= 2\,P(Z > |2.5|) = 2\,(1 - \Phi(2.5)) \\ &= 2\,(1 - 0.9938) \\ &= 2 \times 0.0062 = \boxed{0.0124}\end{aligned}$$

Decision for $\alpha = 0.05$: P-value $= 0.0124 < 0.05 = \alpha \quad \Rightarrow$ **Reject** H_0.

How Small is "Small Enough"?

See Table 4.4 in the textbook

The P-value tells us the **strength of evidence <u>against</u>** H_0:

How Small is "Small Enough"?

See Table 4.4 in the textbook

The P-value tells us the **strength of evidence against** H_0:

P-value	**Evidence against H_0**
> 0.10	Little or no evidence
0.05 to 0.10	Weak evidence
0.01 to 0.05	Moderate evidence
0.001 to 0.01	Strong evidence
< 0.001	Very strong evidence

Section 4.6

The Seven-Step Procedure for Hypothesis Tests

Procedure 4.2: The Seven Steps of a Hypothesis Test

1. **Parameter:** Identify the parameter of interest (μ, σ^2, or p)

Procedure 4.2: The Seven Steps of a Hypothesis Test

1. **Parameter:** Identify the parameter of interest (μ, σ^2, or p)
2. **Null hypothesis:** State H_0

Procedure 4.2: The Seven Steps of a Hypothesis Test

1. **Parameter:** Identify the parameter of interest (μ, σ^2, or p)
2. **Null hypothesis:** State H_0
3. **Alternative hypothesis:** State H_1 (this fixes the tail direction)

Procedure 4.2: The Seven Steps of a Hypothesis Test

1. **Parameter:** Identify the parameter of interest (μ, σ^2, or p)
2. **Null hypothesis:** State H_0
3. **Alternative hypothesis:** State H_1 (this fixes the tail direction)
4. **Test statistic:** Choose it to match the parameter and conditions

Procedure 4.2: The Seven Steps of a Hypothesis Test

1. **Parameter:** Identify the parameter of interest (μ, σ^2, or p)
2. **Null hypothesis:** State H_0
3. **Alternative hypothesis:** State H_1 (this fixes the tail direction)
4. **Test statistic:** Choose it to match the parameter and conditions
5. **Reject H_0 if:** State the rejection criterion for the chosen α (or plan the P-value comparison, Procedure 4.1)

Procedure 4.2: The Seven Steps of a Hypothesis Test

1. **Parameter:** Identify the parameter of interest (μ, σ^2, or p)
2. **Null hypothesis:** State H_0
3. **Alternative hypothesis:** State H_1 (this fixes the tail direction)
4. **Test statistic:** Choose it to match the parameter and conditions
5. **Reject H_0 if:** State the rejection criterion for the chosen α (or plan the P-value comparison, Procedure 4.1)
6. **Computations:** Substitute the sample numbers and compute

Procedure 4.2: The Seven Steps of a Hypothesis Test

1. **Parameter:** Identify the parameter of interest (μ, σ^2, or p)
2. **Null hypothesis:** State H_0
3. **Alternative hypothesis:** State H_1 (this fixes the tail direction)
4. **Test statistic:** Choose it to match the parameter and conditions
5. **Reject H_0 if:** State the rejection criterion for the chosen α (or plan the P-value comparison, Procedure 4.1)
6. **Computations:** Substitute the sample numbers and compute
7. **Conclusion:** Decide, and state the conclusion **in context**

Procedure 4.2: The Seven Steps of a Hypothesis Test

1. **Parameter:** Identify the parameter of interest (μ, σ^2, or p)
2. **Null hypothesis:** State H_0
3. **Alternative hypothesis:** State H_1 (this fixes the tail direction)
4. **Test statistic:** Choose it to match the parameter and conditions
5. **Reject H_0 if:** State the rejection criterion for the chosen α (or plan the P-value comparison, Procedure 4.1)
6. **Computations:** Substitute the sample numbers and compute
7. **Conclusion:** Decide, and state the conclusion **in context**

Always state the conclusion in context of the problem
"We reject H_0: there is sufficient evidence at the 5% significance level that the true mean dwell time is not 71 hours."

Textbook: Chapter 4, §4.6

Variations of Tests

Z-test for the mean (σ known)
t-test for the mean (σ unknown)

Z-Test for the Mean (σ Known)

Test statistic for the mean μ with σ known:

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

Z-Test for the Mean (σ Known)

Test statistic for the mean μ with σ known:

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

Later in Chapter 4: the same Z idea also tests a proportion p — see (4.8) (Lecture 10).

Z-Test for the Mean (σ Known)

Test statistic for the mean μ with σ known:

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

Later in Chapter 4: the same Z idea also tests a proportion p — see (4.8) (Lecture 10).

Assumption: Population is Normal, or the CLT applies (what conditions?)

Z-Test Decision Rule: Visual Guide

Decision Rule:

Z-Test Decision Rule: Visual Guide

Decision Rule:

- **Two-tailed test** ($H_1 : \mu \neq \mu_0$): Reject H_0 if $|Z_0| > z_{\alpha/2}$

Z-Test Decision Rule: Visual Guide

Decision Rule:

- **Two-tailed test** ($H_1 : \mu \neq \mu_0$): Reject H_0 if $|Z_0| > z_{\alpha/2}$
- **Right-tailed test** ($H_1 : \mu > \mu_0$): Reject H_0 if $Z_0 > z_\alpha$

Z-Test Decision Rule: Visual Guide

Decision Rule:

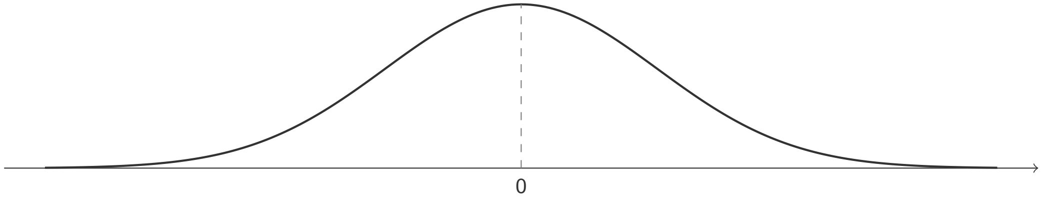
- **Two-tailed test** ($H_1 : \mu \neq \mu_0$): Reject H_0 if $|Z_0| > z_{\alpha/2}$
- **Right-tailed test** ($H_1 : \mu > \mu_0$): Reject H_0 if $Z_0 > z_\alpha$
- **Left-tailed test** ($H_1 : \mu < \mu_0$): Reject H_0 if $Z_0 < -z_\alpha$

Z-Test Decision Rule: Visual Guide

Decision Rule:

- **Two-tailed test** ($H_1 : \mu \neq \mu_0$): Reject H_0 if $|Z_0| > z_{\alpha/2}$
- **Right-tailed test** ($H_1 : \mu > \mu_0$): Reject H_0 if $Z_0 > z_\alpha$
- **Left-tailed test** ($H_1 : \mu < \mu_0$): Reject H_0 if $Z_0 < -z_\alpha$

The **reject** direction in a one-sided test depends on the H_1 sign.

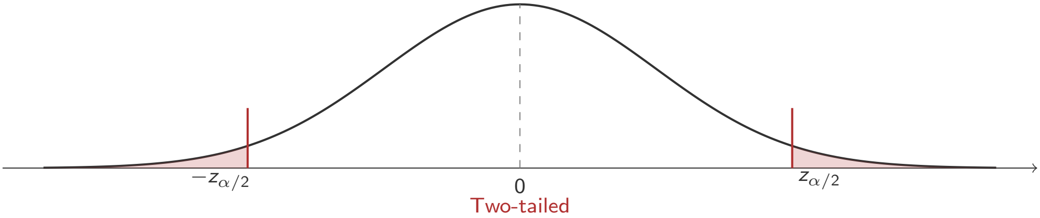


Z-Test Decision Rule: Visual Guide

Decision Rule:

- **Two-tailed test** ($H_1 : \mu \neq \mu_0$): Reject H_0 if $|Z_0| > z_{\alpha/2}$
- **Right-tailed test** ($H_1 : \mu > \mu_0$): Reject H_0 if $Z_0 > z_\alpha$
- **Left-tailed test** ($H_1 : \mu < \mu_0$): Reject H_0 if $Z_0 < -z_\alpha$

The **reject** direction in a one-sided test depends on the H_1 sign.

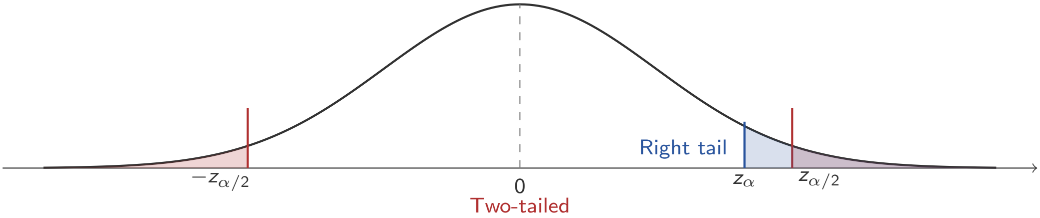


Z-Test Decision Rule: Visual Guide

Decision Rule:

- **Two-tailed test** ($H_1 : \mu \neq \mu_0$): Reject H_0 if $|Z_0| > z_{\alpha/2}$
- **Right-tailed test** ($H_1 : \mu > \mu_0$): Reject H_0 if $Z_0 > z_\alpha$
- **Left-tailed test** ($H_1 : \mu < \mu_0$): Reject H_0 if $Z_0 < -z_\alpha$

The **reject** direction in a one-sided test depends on the H_1 sign.

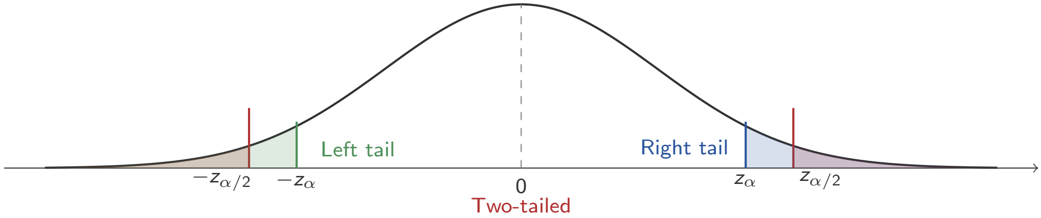


Z-Test Decision Rule: Visual Guide

Decision Rule:

- **Two-tailed test** ($H_1 : \mu \neq \mu_0$): Reject H_0 if $|Z_0| > z_{\alpha/2}$
- **Right-tailed test** ($H_1 : \mu > \mu_0$): Reject H_0 if $Z_0 > z_\alpha$
- **Left-tailed test** ($H_1 : \mu < \mu_0$): Reject H_0 if $Z_0 < -z_\alpha$

The **reject** direction in a one-sided test depends on the H_1 sign.



A startup claims its redesigned onboarding raises mean session length **above** 200 seconds. A sample of $n = 36$ user sessions gives $\bar{x} = 206$ s. Assume $\sigma = 18$ s. Test at $\alpha = 0.05$.

Walking Procedure 4.2

A startup claims its redesigned onboarding raises mean session length **above** 200 seconds. A sample of $n = 36$ user sessions gives $\bar{x} = 206$ s. Assume $\sigma = 18$ s. Test at $\alpha = 0.05$.

Walking Procedure 4.2

Steps 1–3: Parameter: μ (mean session length), σ known.

A startup claims its redesigned onboarding raises mean session length **above** 200 seconds. A sample of $n = 36$ user sessions gives $\bar{x} = 206$ s. Assume $\sigma = 18$ s. Test at $\alpha = 0.05$.

Walking Procedure 4.2

Steps 1–3: Parameter: μ (mean session length), σ known.

- $H_0 : \mu = 200$ vs $H_1 : \mu > 200$ (right-tailed), $\alpha = 0.05$

A startup claims its redesigned onboarding raises mean session length **above** 200 seconds. A sample of $n = 36$ user sessions gives $\bar{x} = 206$ s. Assume $\sigma = 18$ s. Test at $\alpha = 0.05$.

Walking Procedure 4.2

Steps 1–3: Parameter: μ (mean session length), σ known.

- $H_0 : \mu = 200$ vs $H_1 : \mu > 200$ (right-tailed), $\alpha = 0.05$

Steps 4–5: Z-test (4.3); reject if $Z_0 > z_{0.05} = 1.645$

A startup claims its redesigned onboarding raises mean session length **above** 200 seconds. A sample of $n = 36$ user sessions gives $\bar{x} = 206$ s. Assume $\sigma = 18$ s. Test at $\alpha = 0.05$.

Walking Procedure 4.2

Steps 1–3: Parameter: μ (mean session length), σ known.

- $H_0 : \mu = 200$ vs $H_1 : \mu > 200$ (right-tailed), $\alpha = 0.05$

Steps 4–5: Z-test (4.3); reject if $Z_0 > z_{0.05} = 1.645$

Step 6 (Computations): $Z_0 = \dfrac{206 - 200}{18/\sqrt{36}} = \dfrac{6}{3} = 2.0$

A startup claims its redesigned onboarding raises mean session length **above** 200 seconds. A sample of $n = 36$ user sessions gives $\bar{x} = 206$ s. Assume $\sigma = 18$ s. Test at $\alpha = 0.05$.

Walking Procedure 4.2

Steps 1–3: Parameter: μ (mean session length), σ known.

- $H_0 : \mu = 200$ vs $H_1 : \mu > 200$ (right-tailed), $\alpha = 0.05$

Steps 4–5: Z-test (4.3); reject if $Z_0 > z_{0.05} = 1.645$

Step 6 (Computations): $Z_0 = \dfrac{206 - 200}{18/\sqrt{36}} = \dfrac{6}{3} = 2.0$

Check with a P-value (4.5): $P(Z > 2.0) = 1 - \Phi(2.0) = 1 - 0.9772 = 0.0228 < 0.05$

A startup claims its redesigned onboarding raises mean session length **above** 200 seconds. A sample of $n = 36$ user sessions gives $\bar{x} = 206$ s. Assume $\sigma = 18$ s. Test at $\alpha = 0.05$.

Walking Procedure 4.2

Steps 1–3: Parameter: μ (mean session length), σ known.

- $H_0 : \mu = 200$ vs $H_1 : \mu > 200$ (right-tailed), $\alpha = 0.05$

Steps 4–5: Z-test (4.3); reject if $Z_0 > z_{0.05} = 1.645$

Step 6 (Computations): $Z_0 = \dfrac{206 - 200}{18/\sqrt{36}} = \dfrac{6}{3} = 2.0$

Check with a P-value (4.5): $P(Z > 2.0) = 1 - \Phi(2.0) = 1 - 0.9772 = 0.0228 < 0.05$

Step 7 (Conclusion): Since $Z_0 = 2.0 > 1.645$, **reject** H_0. Sufficient evidence at 5% that the mean session length exceeds 200 seconds.

Test for the Mean μ When σ is *Unknown* (t-test)

Scenario: Sample size n, sample mean $\bar{x}$, and sample standard deviation s. σ is unknown, data is approximately normal.

Test for the Mean μ When σ is *Unknown* (t-test)

Scenario: Sample size n, sample mean $\bar{x}$, and sample standard deviation s. σ is unknown, data is approximately normal.

Test statistic:

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{4.6}$$

where T_0 follows a t-distribution with $n-1$ degrees of freedom under H_0.

Test for the Mean μ When σ is *Unknown* (t-test)

Scenario: Sample size n, sample mean $\bar{x}$, and sample standard deviation s. σ is unknown, data is approximately normal.

Test statistic:

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{4.6}$$

where T_0 follows a t-distribution with $n-1$ degrees of freedom under H_0.

Decision rule: Reject H_0 if $|t_0| > t_{\alpha/2,\, n-1}$ (two-tailed); one-tailed rules mirror the Z-test.

Test for the Mean μ When σ is *Unknown* (t-test)

Scenario: Sample size n, sample mean $\bar{x}$, and sample standard deviation s. σ is unknown, data is approximately normal.

Test statistic:

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{4.6}$$

where T_0 follows a t-distribution with $n-1$ degrees of freedom under H_0.

Decision rule: Reject H_0 if $|t_0| > t_{\alpha/2,\, n-1}$ (two-tailed); one-tailed rules mirror the Z-test.

Same swap as Chapter 3: interval (3.1) became (3.5); test (4.3) becomes (4.6).

A vendor claims its fine-tuned guardrail model scores **above** 12.5 (out of 20) on a safety benchmark. Data: $n = 9$ eval suites, $\bar{x} = 13.2$, $s = 1.5$. Test at $\alpha = 0.05$.

Walking Procedure 4.2 again

A vendor claims its fine-tuned guardrail model scores **above** 12.5 (out of 20) on a safety benchmark. Data: $n = 9$ eval suites, $\bar{x} = 13.2$, $s = 1.5$. Test at $\alpha = 0.05$.

Walking Procedure 4.2 again

Steps 1–3: Parameter: μ (mean benchmark score), σ unknown.

A vendor claims its fine-tuned guardrail model scores **above** 12.5 (out of 20) on a safety benchmark. Data: $n = 9$ eval suites, $\bar{x} = 13.2$, $s = 1.5$. Test at $\alpha = 0.05$.

Walking Procedure 4.2 again

Steps 1–3: Parameter: μ (mean benchmark score), σ unknown.

- $H_0 : \mu = 12.5$ vs $H_1 : \mu > 12.5$ (right-tailed), $\alpha = 0.05$

A vendor claims its fine-tuned guardrail model scores **above** 12.5 (out of 20) on a safety benchmark. Data: $n = 9$ eval suites, $\bar{x} = 13.2$, $s = 1.5$. Test at $\alpha = 0.05$.

Walking Procedure 4.2 again

Steps 1–3: Parameter: μ (mean benchmark score), σ unknown.

- $H_0 : \mu = 12.5$ vs $H_1 : \mu > 12.5$ (right-tailed), $\alpha = 0.05$

Steps 4–5: t-test (4.6), df $= 8$; reject if $t_0 > t_{0.05,8} = 1.860$

A vendor claims its fine-tuned guardrail model scores **above** 12.5 (out of 20) on a safety benchmark. Data: $n = 9$ eval suites, $\bar{x} = 13.2$, $s = 1.5$. Test at $\alpha = 0.05$.

Walking Procedure 4.2 again

Steps 1–3: Parameter: μ (mean benchmark score), σ unknown.

- $H_0 : \mu = 12.5$ vs $H_1 : \mu > 12.5$ (right-tailed), $\alpha = 0.05$

Steps 4–5: t-test (4.6), df $= 8$; reject if $t_0 > t_{0.05,\,8} = 1.860$

Step 6 (Computations): $t_0 = \dfrac{13.2 - 12.5}{1.5/\sqrt{9}} = \dfrac{0.7}{0.5} = 1.4$

A vendor claims its fine-tuned guardrail model scores **above** 12.5 (out of 20) on a safety benchmark. Data: $n = 9$ eval suites, $\bar{x} = 13.2$, $s = 1.5$. Test at $\alpha = 0.05$.

Walking Procedure 4.2 again

Steps 1–3: Parameter: μ (mean benchmark score), σ unknown.

- $H_0 : \mu = 12.5$ vs $H_1 : \mu > 12.5$ (right-tailed), $\alpha = 0.05$

Steps 4–5: t-test (4.6), df $= 8$; reject if $t_0 > t_{0.05,8} = 1.860$

Step 6 (Computations): $t_0 = \dfrac{13.2 - 12.5}{1.5/\sqrt{9}} = \dfrac{0.7}{0.5} = 1.4$

P-value bounds (Procedure 4.3): from the t-table, $t_{0.10,8} = 1.397 < 1.4 < 1.860 = t_{0.05,8}$, so $0.05 < \text{P-value} < 0.10$

A vendor claims its fine-tuned guardrail model scores **above** 12.5 (out of 20) on a safety benchmark. Data: $n = 9$ eval suites, $\bar{x} = 13.2$, $s = 1.5$. Test at $\alpha = 0.05$.

Walking Procedure 4.2 again

Steps 1–3: Parameter: μ (mean benchmark score), σ unknown.

- $H_0 : \mu = 12.5$ vs $H_1 : \mu > 12.5$ (right-tailed), $\alpha = 0.05$

Steps 4–5: t-test (4.6), df $= 8$; reject if $t_0 > t_{0.05,8} = 1.860$

Step 6 (Computations): $t_0 = \dfrac{13.2 - 12.5}{1.5/\sqrt{9}} = \dfrac{0.7}{0.5} = 1.4$

P-value bounds (Procedure 4.3): from the t-table, $t_{0.10,8} = 1.397 < 1.4 < 1.860 = t_{0.05,8}$, so $0.05 <$ P-value < 0.10

Step 7 (Conclusion): Since $t_0 = 1.4 < 1.860$, **fail to reject H_0**. **Not enough evidence** at 5% that the mean benchmark score exceeds 12.5.

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1
  - Test statistic and rejection criterion

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1
  - Test statistic and rejection criterion
  - Computations

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1
  - Test statistic and rejection criterion
  - Computations
  - Conclusion in context

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1
  - Test statistic and rejection criterion
  - Computations
  - Conclusion in context
- Types of tests:

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1
  - Test statistic and rejection criterion
  - Computations
  - Conclusion in context
- Types of tests:
  - Z-test (4.3): mean μ with σ known

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1
  - Test statistic and rejection criterion
  - Computations
  - Conclusion in context
- Types of tests:
  - Z-test (4.3): mean μ with σ known
  - t-test (4.6): mean μ with σ unknown

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1
  - Test statistic and rejection criterion
  - Computations
  - Conclusion in context
- Types of tests:
  - Z-test (4.3): mean μ with σ known
  - t-test (4.6): mean μ with σ unknown
  - Coming next lecture:
    - χ^2-test for the variance (4.7) and Z-test for a proportion (4.8)

Lecture 9 Summary

- **P-value** (4.4): smallest α at which we'd reject; compare to α via Procedure 4.1
- **Seven steps** (Procedure 4.2):
  - Parameter, H_0, H_1
  - Test statistic and rejection criterion
  - Computations
  - Conclusion in context
- Types of tests:
  - Z-test (4.3): mean μ with σ known
  - t-test (4.6): mean μ with σ unknown
  - Coming next lecture:
    - χ^2-test for the variance (4.7) and Z-test for a proportion (4.8)
    - Type II error (β) and the α–β tradeoff