# ISE 315: Engineering Statistics

*Lecture 11: Statistical Inference for Two Samples*

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 11

Statistical Inference for Two Samples (Chapter 5, Part 1)

## Lecture 11 Outline

- Announcements:
  - HW #4 due next week; Quiz 4 (Chapter 5) before class next week
  - Major Exam 1 in Week 8 (Chapters 2–5)

## Lecture 11 Outline

- Announcements:
  - HW #4 due next week; Quiz 4 (Chapter 5) before class next week
  - Major Exam 1 in Week 8 (Chapters 2–5)
- From one sample to two samples

# Lecture 11 Outline

- Announcements:
  - HW #4 due next week; Quiz 4 (Chapter 5) before class next week
  - Major Exam 1 in Week 8 (Chapters 2–5)
- From one sample to two samples
- Difference in means $\mu_1 - \mu_2$, variances **known**: two-sample $Z$

## Lecture 11 Outline

- Announcements:
  - HW #4 due next week; Quiz 4 (Chapter 5) before class next week
  - Major Exam 1 in Week 8 (Chapters 2–5)
- From one sample to two samples
- Difference in means $\mu_1 - \mu_2$, variances **known**: two-sample $Z$
- Difference in means, variances **unknown but equal**: pooled $t$

## Lecture 11 Outline

- Announcements:
  - HW #4 due next week; Quiz 4 (Chapter 5) before class next week
  - Major Exam 1 in Week 8 (Chapters 2–5)
- From one sample to two samples
- Difference in means $\mu_1 - \mu_2$, variances **known**: two-sample $Z$
- Difference in means, variances **unknown but equal**: pooled $t$
- Worked examples and a quick practice

# From One Sample to Two Samples

*See Figure 5.1 in the textbook*

$$\begin{array}{c} \text{Population} \\ \mu,\ \sigma^2 \end{array} \longrightarrow (n){:}\ \bar{X},\ S^2$$

**Ch. 3–4: One population**

$H_0: \mu = \mu_0$ or $H_0: \sigma^2 = \sigma_0^2$

Textbook: Chapter 5, §5.1 ((5.1) onward)

# From One Sample to Two Samples

*See Figure 5.1 in the textbook*

Population

$\mu, \ \sigma^2$

$\downarrow$

$(n)$: $\bar{X}, \ S^2$

**Ch. 3–4: One population**

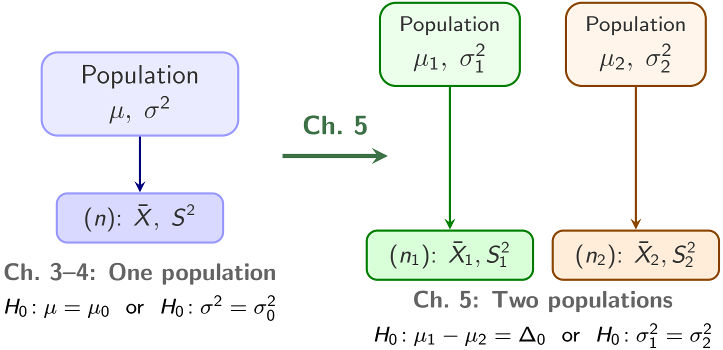
$H_0 : \mu = \mu_0$ or $H_0 : \sigma^2 = \sigma_0^2$

**Ch. 5** $\longrightarrow$

Textbook: Chapter 5, §5.1 ((5.1) onward)

# From One Sample to Two Samples

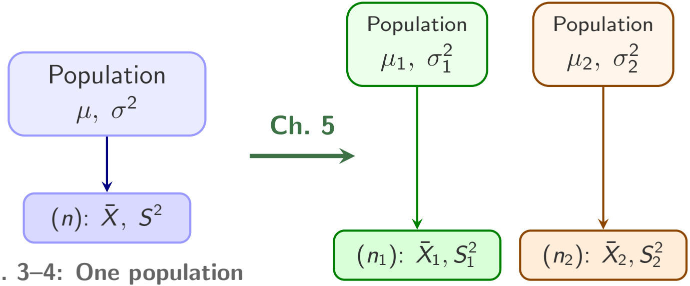
*See Figure 5.1 in the textbook*




Textbook: Chapter 5, §5.1 ((5.1) onward)

# From One Sample to Two Samples

*See Figure 5.1 in the textbook*




**Ch. 3–4: One population**

$H_0: \mu = \mu_0$ or $H_0: \sigma^2 = \sigma_0^2$

**Ch. 5: Two populations**

$H_0: \mu_1 - \mu_2 = \Delta_0$ or $H_0: \sigma_1^2 = \sigma_2^2$

**From Lecture 4** (2.13)–(2.14):

$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$

$\mathrm{Var}[\bar{X}_1 - \bar{X}_2] = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

Textbook: Chapter 5, §5.1 ((5.1) onward)

# Setting Up the Two-Sample Problem

*Difference in Means, $\sigma_1^2$ and $\sigma_2^2$ Known*

**Setup:** Two independent random samples:

- Sample 1: $n_1$ observations from a population with mean $\mu_1$, known variance $\sigma_1^2$
- Sample 2: $n_2$ observations from a population with mean $\mu_2$, known variance $\sigma_2^2$

# Setting Up the Two-Sample Problem

*Difference in Means, $\sigma_1^2$ and $\sigma_2^2$ Known*

---

**Setup:** Two independent random samples:

- Sample 1: $n_1$ observations from a population with mean $\mu_1$, known variance $\sigma_1^2$
- Sample 2: $n_2$ observations from a population with mean $\mu_2$, known variance $\sigma_2^2$

**Point estimator:** The natural estimator of $\mu_1 - \mu_2$ is $\bar{X}_1 - \bar{X}_2$

# Setting Up the Two-Sample Problem

*Difference in Means, $\sigma_1^2$ and $\sigma_2^2$ Known*

---

**Setup:** Two independent random samples:

- Sample 1: $n_1$ observations from a population with mean $\mu_1$, known variance $\sigma_1^2$
- Sample 2: $n_2$ observations from a population with mean $\mu_2$, known variance $\sigma_2^2$

**Point estimator:** The natural estimator of $\mu_1 - \mu_2$ is $\bar{X}_1 - \bar{X}_2$

**Distribution** (Chapter 2, (2.13)):

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

## Setting Up the Two-Sample Problem

*Difference in Means, $\sigma_1^2$ and $\sigma_2^2$ Known*

**Setup:** Two independent random samples:

- Sample 1: $n_1$ observations from a population with mean $\mu_1$, known variance $\sigma_1^2$
- Sample 2: $n_2$ observations from a population with mean $\mu_2$, known variance $\sigma_2^2$

**Point estimator:** The natural estimator of $\mu_1 - \mu_2$ is $\bar{X}_1 - \bar{X}_2$

**Distribution** (Chapter 2, (2.13)):

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

# Hypothesis Test for $\mu_1 - \mu_2$ ($\sigma_1^2, \sigma_2^2$ Known)

**Hypotheses** (testing whether the difference equals some value $\Delta_0$, often 0):

$$H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq \Delta_0$$

## Hypothesis Test for $\mu_1 - \mu_2$ ($\sigma_1^2, \sigma_2^2$ Known)

**Hypotheses** (testing whether the difference equals some value $\Delta_0$, often 0):

$$H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq \Delta_0$$

**Test statistic:**

$$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{5.1}$$

## Hypothesis Test for $\mu_1 - \mu_2$ ($\sigma_1^2, \sigma_2^2$ Known)

**Hypotheses** (testing whether the difference equals some value $\Delta_0$, often 0):

$$H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq \Delta_0$$

**Test statistic:**

$$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{5.1}$$

Under $H_0$, $Z_0 \sim N(0, 1)$.

Decision: reject $H_0$ if the test statistic is beyond the critical value, **or** if P-value $\leq \alpha$ (Procedure 4.1) — same two routes as Chapter 4.

# Decision Rules for the Two-Sample $Z$-Test

*See Table 5.1 in the textbook*

| **Alternative $H_1$** | **Reject $H_0$ if** | **P-value** |
| --- | --- | --- |
| $\mu_1 - \mu_2 \neq \Delta_0$ | $\lvert Z_0 \rvert > z_{\alpha/2}$ | $2\left[1 - \Phi(\lvert z_0 \rvert)\right]$ |

# Decision Rules for the Two-Sample $Z$-Test

*See Table 5.1 in the textbook*

| **Alternative $H_1$** | **Reject $H_0$ if** | **P-value** |
|---|---|---|
| $\mu_1 - \mu_2 \neq \Delta_0$ | $\lvert Z_0 \rvert > z_{\alpha/2}$ | $2\left[1 - \Phi(\lvert z_0 \rvert)\right]$ |
| $\mu_1 - \mu_2 > \Delta_0$ | $Z_0 > z_\alpha$ | $1 - \Phi(z_0)$ |

# Decision Rules for the Two-Sample $Z$-Test

*See Table 5.1 in the textbook*

| **Alternative** $H_1$ | **Reject** $H_0$ **if** | **P-value** |
|---|---|---|
| $\mu_1 - \mu_2 \neq \Delta_0$ | $\lvert Z_0 \rvert > z_{\alpha/2}$ | $2\left[1 - \Phi(\lvert z_0 \rvert)\right]$ |
| $\mu_1 - \mu_2 > \Delta_0$ | $Z_0 > z_\alpha$ | $1 - \Phi(z_0)$ |
| $\mu_1 - \mu_2 < \Delta_0$ | $Z_0 < -z_\alpha$ | $\Phi(z_0)$ |

# Decision Rules for the Two-Sample $Z$-Test

*See Table 5.1 in the textbook*

| **Alternative** $H_1$ | **Reject** $H_0$ **if** | **P-value** |
|---|---|---|
| $\mu_1 - \mu_2 \neq \Delta_0$ | $\lvert Z_0 \rvert > z_{\alpha/2}$ | $2\left[1 - \Phi(\lvert z_0 \rvert)\right]$ |
| $\mu_1 - \mu_2 > \Delta_0$ | $Z_0 > z_\alpha$ | $1 - \Phi(z_0)$ |
| $\mu_1 - \mu_2 < \Delta_0$ | $Z_0 < -z_\alpha$ | $\Phi(z_0)$ |

**Note:** Structurally identical to the single-sample $Z$-test (4.3). Only the computation of $Z_0$ changes.

## Confidence Interval for $\mu_1 - \mu_2$ ($\sigma_1^2, \sigma_2^2$ Known)

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \; \pm \; z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{5.2}$$

## Confidence Interval for $\mu_1 - \mu_2$ ($\sigma_1^2, \sigma_2^2$ Known)

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \ \pm \ z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{5.2}$$

Compare to the single-sample case (3.1):

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

## Confidence Interval for $\mu_1 - \mu_2$ ($\sigma_1^2, \sigma_2^2$ Known)

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \ \pm \ z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{5.2}$$

Compare to the single-sample case (3.1):

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

Same structure: estimate $\pm$ (critical value) $\times$ (standard error). Can you summarize the key differences?

# Example: Comparing Dwell Times at Two Terminals

*Example 5.1 in the textbook*

A logistics analyst compares mean container dwell times (hours) at two port terminals. Long operating histories give known standard deviations.

- Terminal 1: $n_1 = 40$, $\bar{x}_1 = 68.2$ h, $\sigma_1 = 9$ h (known)
- Terminal 2: $n_2 = 50$, $\bar{x}_2 = 72.5$ h, $\sigma_2 = 10$ h (known)

Is there evidence that the mean dwell times differ? Use $\alpha = 0.05$.

## Example: Comparing Dwell Times at Two Terminals

*Example 5.1 in the textbook*

---

A logistics analyst compares mean container dwell times (hours) at two port terminals. Long operating histories give known standard deviations.

- Terminal 1: $n_1 = 40$, $\bar{x}_1 = 68.2$ h, $\sigma_1 = 9$ h (known)
- Terminal 2: $n_2 = 50$, $\bar{x}_2 = 72.5$ h, $\sigma_2 = 10$ h (known)

Is there evidence that the mean dwell times differ? Use $\alpha = 0.05$.

**Steps 1–3 (Procedure 4.2):** Parameter: $\mu_1 - \mu_2$. Both $\sigma$'s known. Samples independent. Use the two-sample $Z$-test (5.1).

# Example: Comparing Dwell Times at Two Terminals

*Example 5.1 in the textbook*

---

A logistics analyst compares mean container dwell times (hours) at two port terminals. Long operating histories give known standard deviations.

- Terminal 1: $n_1 = 40$, $\bar{x}_1 = 68.2$ h, $\sigma_1 = 9$ h (known)
- Terminal 2: $n_2 = 50$, $\bar{x}_2 = 72.5$ h, $\sigma_2 = 10$ h (known)

Is there evidence that the mean dwell times differ? Use $\alpha = 0.05$.

**Steps 1–3 (Procedure 4.2):** Parameter: $\mu_1 - \mu_2$. Both $\sigma$'s known. Samples independent. Use the two-sample $Z$-test (5.1).

$H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 \neq 0$, $\alpha = 0.05$.

## Example: Terminal Dwell Times (Solution)

**Steps 4–5:** Test statistic (5.1); two-tailed, $z_{0.025} = 1.96$; reject if $|Z_0| > 1.96$.

# Example: Terminal Dwell Times (Solution)

**Steps 4–5:** Test statistic (5.1); two-tailed, $z_{0.025} = 1.96$; reject if $|Z_0| > 1.96$.

**Step 6 (Computations):**

$$Z_0 = \frac{(68.2 - 72.5) - 0}{\sqrt{\dfrac{9^2}{40} + \dfrac{10^2}{50}}}$$

## Example: Terminal Dwell Times (Solution)

**Steps 4–5:** Test statistic (5.1); two-tailed, $z_{0.025} = 1.96$; reject if $|Z_0| > 1.96$.

**Step 6 (Computations):**

$$Z_0 = \frac{(68.2 - 72.5) - 0}{\sqrt{\frac{9^2}{40} + \frac{10^2}{50}}} = \frac{-4.3}{\sqrt{2.025 + 2.000}}$$

## Example: Terminal Dwell Times (Solution)

**Steps 4–5:** Test statistic (5.1); two-tailed, $z_{0.025} = 1.96$; reject if $|Z_0| > 1.96$.

**Step 6 (Computations):**

$$Z_0 = \frac{(68.2 - 72.5) - 0}{\sqrt{\dfrac{9^2}{40} + \dfrac{10^2}{50}}} = \frac{-4.3}{\sqrt{2.025 + 2.000}} = \frac{-4.3}{2.006} = -2.143$$

## Example: Terminal Dwell Times (Solution)

**Steps 4–5:** Test statistic (5.1); two-tailed, $z_{0.025} = 1.96$; reject if $|Z_0| > 1.96$.

**Step 6 (Computations):**

$$Z_0 = \frac{(68.2 - 72.5) - 0}{\sqrt{\dfrac{9^2}{40} + \dfrac{10^2}{50}}} = \frac{-4.3}{\sqrt{2.025 + 2.000}} = \frac{-4.3}{2.006} = -2.143$$

**P-value:** $2\,[1 - \Phi(2.14)] = 2\,(1 - 0.9838) = 0.0324$

## Example: Terminal Dwell Times (Solution)

**Steps 4–5:** Test statistic (5.1); two-tailed, $z_{0.025} = 1.96$; reject if $|Z_0| > 1.96$.

**Step 6 (Computations):**

$$Z_0 = \frac{(68.2 - 72.5) - 0}{\sqrt{\frac{9^2}{40} + \frac{10^2}{50}}} = \frac{-4.3}{\sqrt{2.025 + 2.000}} = \frac{-4.3}{2.006} = -2.143$$

**P-value:** $2\,[1 - \Phi(2.14)] = 2\,(1 - 0.9838) = 0.0324$

**Step 7 (Conclusion):** $|Z_0| = 2.143 > 1.96$ (and $0.0324 < 0.05$). **Reject $H_0$.** There is sufficient evidence at the 5% level that the mean dwell times at the two terminals differ.

## Example: Terminal Dwell Times (Confidence Interval)

**95% CI for** $\mu_1 - \mu_2$ via (5.2):

$$(\bar{x}_1 - \bar{x}_2) \;\pm\; z_{0.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

# Example: Terminal Dwell Times (Confidence Interval)

**95% CI for** $\mu_1 - \mu_2$ via (5.2):

$$(\bar{x}_1 - \bar{x}_2) \; \pm \; z_{0.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = -4.3 \pm 1.96\,(2.006)$$

# Example: Terminal Dwell Times (Confidence Interval)

**95% CI for** $\mu_1 - \mu_2$ via (5.2):

$$
\begin{aligned}
(\bar{x}_1 - \bar{x}_2) \;\pm\; z_{0.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= -4.3 \pm 1.96\,(2.006) \\
&= -4.3 \pm 3.932
\end{aligned}
$$

## Example: Terminal Dwell Times (Confidence Interval)

**95% CI for** $\mu_1 - \mu_2$ via (5.2):

$$
\begin{aligned}
(\bar{x}_1 - \bar{x}_2) \;\pm\; z_{0.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= -4.3 \pm 1.96\,(2.006) \\
&= -4.3 \pm 3.932 \\
&= (-8.232,\; -0.368) \text{ hours}
\end{aligned}
$$

## Example: Terminal Dwell Times (Confidence Interval)

**95% CI for** $\mu_1 - \mu_2$ via (5.2):

$$\begin{aligned}(\bar{x}_1 - \bar{x}_2) \;\pm\; z_{0.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= -4.3 \pm 1.96\,(2.006) \\ &= -4.3 \pm 3.932 \\ &= (-8.232,\; -0.368) \text{ hours}\end{aligned}$$

The interval **excludes** 0, consistent with rejecting $H_0$ — the CI–test duality from Chapter 4.

## Example: Terminal Dwell Times (Confidence Interval)

**95% CI for** $\mu_1 - \mu_2$ via (5.2):

$$
\begin{aligned}
(\bar{x}_1 - \bar{x}_2) \;\pm\; z_{0.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= -4.3 \pm 1.96\,(2.006) \\
&= -4.3 \pm 3.932 \\
&= (-8.232,\; -0.368) \text{ hours}
\end{aligned}
$$

The interval **excludes** 0, consistent with rejecting $H_0$ — the CI–test duality from Chapter 4.

**Interpretation:** We are 95% confident Terminal 1's mean dwell time is between 0.37 and 8.23 hours *shorter* than Terminal 2's.

# Two Cases When Variances Are Unknown

*Difference in Means, $\sigma_1^2$ and $\sigma_2^2$ Unknown*

When $\sigma_1^2$ and $\sigma_2^2$ are unknown, we estimate them with $s_1^2$ and $s_2^2$.

But we need to decide: can we assume $\sigma_1^2 = \sigma_2^2$?

Textbook: Chapter 5, §5.3 ((5.3) onward)

# Two Cases When Variances Are Unknown

*Difference in Means, $\sigma_1^2$ and $\sigma_2^2$ Unknown*

---

When $\sigma_1^2$ and $\sigma_2^2$ are unknown, we estimate them with $s_1^2$ and $s_2^2$.

But we need to decide: can we assume $\sigma_1^2 = \sigma_2^2$?

**Case 1: $\sigma_1^2 = \sigma_2^2$ (equal variances assumed)**
We "pool" the two sample variances into one combined estimate.
This is the **pooled $t$-test** — §5.3, today.

Textbook: Chapter 5, §5.3 ((5.3) onward)

# Two Cases When Variances Are Unknown

*Difference in Means, $\sigma_1^2$ and $\sigma_2^2$ Unknown*

When $\sigma_1^2$ and $\sigma_2^2$ are unknown, we estimate them with $s_1^2$ and $s_2^2$.

But we need to decide: can we assume $\sigma_1^2 = \sigma_2^2$?

**Case 1: $\sigma_1^2 = \sigma_2^2$ (equal variances assumed)**
We "pool" the two sample variances into one combined estimate.
This is the **pooled $t$-test** — §5.3, today.

**Case 2: $\sigma_1^2 \neq \sigma_2^2$ (unequal variances)**
Keep the variances separate and use an **approximate degrees of freedom**.
This is **Welch's $t$-test** — §5.4, next lecture.

Textbook: Chapter 5, §5.3 ((5.3) onward)

## The Pooled $t$-Test (assuming $\sigma_1^2 = \sigma_2^2$)

**Pooled variance estimator** (a weighted average of the sample variances):

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2} \tag{5.3}$$

## The Pooled $t$-Test (assuming $\sigma_1^2 = \sigma_2^2$)

**Pooled variance estimator** (a weighted average of the sample variances):

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2} \tag{5.3}$$

**Test statistic:**

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{5.4}$$

## The Pooled $t$-Test (assuming $\sigma_1^2 = \sigma_2^2$)

**Pooled variance estimator** (a weighted average of the sample variances):

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2} \tag{5.3}$$

**Test statistic:**

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{5.4}$$

- **Under** $H_0$**,** $T_0 \sim t_{n_1+n_2-2}$**.**

# The Pooled $t$-Test (assuming $\sigma_1^2 = \sigma_2^2$)

**Pooled variance estimator** (a weighted average of the sample variances):

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2} \tag{5.3}$$

**Test statistic:**

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{5.4}$$

- **Under** $H_0$**,** $T_0 \sim t_{n_1+n_2-2}$**.**
- **Degrees of freedom:** $n_1 + n_2 - 2$ (both samples estimate the common variance).

## Confidence Interval: Pooled $t$ ($\sigma_1^2 = \sigma_2^2$)

A $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$:

$$(\bar{x}_1 - \bar{x}_2) \ \pm \ t_{\alpha/2,\, n_1+n_2-2} \ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{5.5}$$

## Confidence Interval: Pooled $t$ ($\sigma_1^2 = \sigma_2^2$)

A $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$:

$$(\bar{x}_1 - \bar{x}_2) \ \pm \ t_{\alpha/2,\, n_1+n_2-2} \ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{5.5}$$

**Compare to the single-sample $t$-CI** (3.5):

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

## Confidence Interval: Pooled $t$ $(\sigma_1^2 = \sigma_2^2)$

A $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$:

$$(\bar{x}_1 - \bar{x}_2) \;\pm\; t_{\alpha/2,\, n_1+n_2-2}\; s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{5.5}$$

**Compare to the single-sample $t$-CI** (3.5):

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

- Same structure: estimate $\pm$ (critical value) $\times$ (standard error)

## Confidence Interval: Pooled $t$ $(\sigma_1^2 = \sigma_2^2)$

A $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$:

$$(\bar{x}_1 - \bar{x}_2) \;\pm\; t_{\alpha/2,\, n_1+n_2-2}\; s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{5.5}$$

**Compare to the single-sample $t$-CI** (3.5):

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

- Same structure: estimate $\pm$ (critical value) $\times$ (standard error)
- The standard error now carries *two* sample sizes and a *pooled* variance

## Decision Rules for the Pooled $t$-Test

Let $\nu = n_1 + n_2 - 2$.

| **Alternative** $H_1$ | **Reject** $H_0$ **if** | **P-value** |
|---|---|---|
| $\mu_1 - \mu_2 \neq \Delta_0$ | $\lvert T_0 \rvert > t_{\alpha/2,\,\nu}$ | $2\,P(T > \lvert t_0 \rvert)$ |

## Decision Rules for the Pooled $t$-Test

Let $\nu = n_1 + n_2 - 2$.

| **Alternative $H_1$** | **Reject $H_0$ if** | **P-value** |
|---|---|---|
| $\mu_1 - \mu_2 \neq \Delta_0$ | $\lvert T_0 \rvert > t_{\alpha/2,\,\nu}$ | $2\,P(T > \lvert t_0 \rvert)$ |
| $\mu_1 - \mu_2 > \Delta_0$ | $T_0 > t_{\alpha,\,\nu}$ | $P(T > t_0)$ |

# Decision Rules for the Pooled $t$-Test

Let $\nu = n_1 + n_2 - 2$.

| **Alternative $H_1$** | **Reject $H_0$ if** | **P-value** |
|---|---|---|
| $\mu_1 - \mu_2 \neq \Delta_0$ | $\lvert T_0 \rvert > t_{\alpha/2,\,\nu}$ | $2\,P(T > \lvert t_0 \rvert)$ |
| $\mu_1 - \mu_2 > \Delta_0$ | $T_0 > t_{\alpha,\,\nu}$ | $P(T > t_0)$ |
| $\mu_1 - \mu_2 < \Delta_0$ | $T_0 < -t_{\alpha,\,\nu}$ | $P(T < t_0)$ |

# Decision Rules for the Pooled $t$-Test

Let $\nu = n_1 + n_2 - 2$.

| **Alternative** $H_1$ | **Reject** $H_0$ **if** | **P-value** |
|---|---|---|
| $\mu_1 - \mu_2 \neq \Delta_0$ | $\lvert T_0 \rvert > t_{\alpha/2,\,\nu}$ | $2\,P(T > \lvert t_0 \rvert)$ |
| $\mu_1 - \mu_2 > \Delta_0$ | $T_0 > t_{\alpha,\,\nu}$ | $P(T > t_0)$ |
| $\mu_1 - \mu_2 < \Delta_0$ | $T_0 < -t_{\alpha,\,\nu}$ | $P(T < t_0)$ |

**Reminder:** Bound P-values from the $t$-table Table B.3 using Procedure 4.3, or use software for exact values.

# Example: A/B Test of Two Onboarding Flows

*Example 5.2 in the textbook*

A startup A/B tests two onboarding flows and records the session length (minutes) of new users. Assume equal population variances.

- Flow A: $n_1 = 12$, $\bar{x}_1 = 18.6$, $s_1 = 2.4$
- Flow B: $n_2 = 14$, $\bar{x}_2 = 16.9$, $s_2 = 2.1$

Do the mean session lengths differ? Test at $\alpha = 0.05$.

## Example: A/B Test of Two Onboarding Flows

*Example 5.2 in the textbook*

---

A startup A/B tests two onboarding flows and records the session length (minutes) of new users. Assume equal population variances.

- Flow A: $n_1 = 12$, $\bar{x}_1 = 18.6$, $s_1 = 2.4$
- Flow B: $n_2 = 14$, $\bar{x}_2 = 16.9$, $s_2 = 2.1$

Do the mean session lengths differ? Test at $\alpha = 0.05$.

**Steps 1–3:** Parameter: $\mu_1 - \mu_2$. Variances unknown but assumed equal ($s_1/s_2 = 1.14 < 2$). Independent samples. Use the pooled $t$-test (5.4).

# Example: A/B Test of Two Onboarding Flows

*Example 5.2 in the textbook*

A startup A/B tests two onboarding flows and records the session length (minutes) of new users. Assume equal population variances.

- Flow A: $n_1 = 12$, $\bar{x}_1 = 18.6$, $s_1 = 2.4$
- Flow B: $n_2 = 14$, $\bar{x}_2 = 16.9$, $s_2 = 2.1$

Do the mean session lengths differ? Test at $\alpha = 0.05$.

**Steps 1–3:** Parameter: $\mu_1 - \mu_2$. Variances unknown but assumed equal ($s_1/s_2 = 1.14 < 2$). Independent samples. Use the pooled $t$-test (5.4).

$H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 \neq 0$, $\alpha = 0.05$.

# Example: Onboarding Flows (Solution)

**Step 6a:** First the pooled variance (5.3):

$$s_p^2 = \frac{(12-1)(2.4)^2 + (14-1)(2.1)^2}{12+14-2}$$

## Example: Onboarding Flows (Solution)

**Step 6a:** First the pooled variance (5.3):

$$s_p^2 = \frac{(12-1)(2.4)^2 + (14-1)(2.1)^2}{12+14-2} = \frac{11(5.76) + 13(4.41)}{24}$$

## Example: Onboarding Flows (Solution)

**Step 6a:** First the pooled variance (5.3):

$$s_p^2 = \frac{(12-1)(2.4)^2 + (14-1)(2.1)^2}{12+14-2} = \frac{11(5.76) + 13(4.41)}{24}$$
$$= \frac{63.36 + 57.33}{24} = \frac{120.69}{24} = 5.029$$

## Example: Onboarding Flows (Solution)

**Step 6a:** First the pooled variance (5.3):

$$
\begin{aligned}
s_p^2 &= \frac{(12-1)(2.4)^2 + (14-1)(2.1)^2}{12+14-2} = \frac{11(5.76) + 13(4.41)}{24} \\
&= \frac{63.36 + 57.33}{24} = \frac{120.69}{24} = 5.029
\end{aligned}
$$

So $s_p = \sqrt{5.029} = 2.243$.

# Example: Onboarding Flows (Solution)

**Step 6a:** First the pooled variance (5.3):

$$s_p^2 = \frac{(12-1)(2.4)^2 + (14-1)(2.1)^2}{12+14-2} = \frac{11(5.76) + 13(4.41)}{24}$$
$$= \frac{63.36 + 57.33}{24} = \frac{120.69}{24} = 5.029$$

So $s_p = \sqrt{5.029} = 2.243$.

**Step 6b:** Test statistic (5.4):

$$T_0 = \frac{(18.6 - 16.9) - 0}{2.243\sqrt{\frac{1}{12} + \frac{1}{14}}} = \frac{1.7}{2.243 \times 0.3934} = \frac{1.7}{0.8822} = 1.927$$

## Example: Onboarding Flows (Solution)

**Step 6a:** First the pooled variance (5.3):

$$s_p^2 = \frac{(12-1)(2.4)^2 + (14-1)(2.1)^2}{12+14-2} = \frac{11(5.76) + 13(4.41)}{24}$$
$$= \frac{63.36 + 57.33}{24} = \frac{120.69}{24} = 5.029$$

So $s_p = \sqrt{5.029} = 2.243$.

**Step 6b:** Test statistic (5.4):

$$T_0 = \frac{(18.6 - 16.9) - 0}{2.243\sqrt{\frac{1}{12} + \frac{1}{14}}} = \frac{1.7}{2.243 \times 0.3934} = \frac{1.7}{0.8822} = 1.927$$

## Example: Onboarding Flows (Conclusion)

**Steps 4–5:** Two-tailed, $\nu = 24$: $t_{0.025,\,24} = 2.064$. Reject if $|T_0| > 2.064$.

# Example: Onboarding Flows (Conclusion)

**Steps 4–5:** Two-tailed, $\nu = 24$: $t_{0.025,\,24} = 2.064$. Reject if $|T_0| > 2.064$.

**Decision:** $|T_0| = 1.927 < 2.064$. Not in the rejection region.

P-value bounds (Procedure 4.3), $\nu = 24$: $t_{0.05} = 1.711$ and $t_{0.025} = 2.064$.
Since $1.711 < 1.927 < 2.064$: one tail is between 0.025 and 0.05,
so the two-tailed P-value satisfies $0.05 < \text{P-value} < 0.10$.

## Example: Onboarding Flows (Conclusion)

**Steps 4–5:** Two-tailed, $\nu = 24$: $t_{0.025,\,24} = 2.064$. Reject if $|T_0| > 2.064$.

**Decision:** $|T_0| = 1.927 < 2.064$. Not in the rejection region.

P-value bounds (Procedure 4.3), $\nu = 24$: $t_{0.05} = 1.711$ and $t_{0.025} = 2.064$.
Since $1.711 < 1.927 < 2.064$: one tail is between 0.025 and 0.05,
so the two-tailed P-value satisfies $0.05 < \text{P-value} < 0.10$.

**Step 7:** **Fail to reject $H_0$.** At $\alpha = 0.05$, not sufficient evidence that the mean session lengths differ between the two flows.

## Example: Onboarding Flows (Conclusion)

**Steps 4–5:** Two-tailed, $\nu = 24$: $t_{0.025,\,24} = 2.064$. Reject if $|T_0| > 2.064$.

**Decision:** $|T_0| = 1.927 < 2.064$. Not in the rejection region.

P-value bounds (Procedure 4.3), $\nu = 24$: $t_{0.05} = 1.711$ and $t_{0.025} = 2.064$.
Since $1.711 < 1.927 < 2.064$: one tail is between 0.025 and 0.05,
so the two-tailed P-value satisfies $0.05 < \text{P-value} < 0.10$.

**Step 7:** **Fail to reject $H_0$.** At $\alpha = 0.05$, not sufficient evidence that the mean session lengths differ between the two flows.

**95% CI** (5.5): $1.7 \pm 2.064\,(0.8822) = 1.7 \pm 1.821 = (-0.121,\ 3.521)$ min.
The interval contains 0, consistent with failing to reject.

## Quick Practice: Which Two-Sample Tool?

*AI safety testing scenarios*

- **(a)** Two AV software builds; disengagement-free distance per intervention, $\sigma_1, \sigma_2$ known from years of fleet logs; large independent samples.

## Quick Practice: Which Two-Sample Tool?

*AI safety testing scenarios*

- **(a)** Two AV software builds; disengagement-free distance per intervention, $\sigma_1, \sigma_2$ known from years of fleet logs; large independent samples.
  $\Rightarrow$ Two-sample $Z$-test (5.1)

# Quick Practice: Which Two-Sample Tool?

*AI safety testing scenarios*

- **(a)** Two AV software builds; disengagement-free distance per intervention, $\sigma_1, \sigma_2$ known from years of fleet logs; large independent samples.
  $\Rightarrow$ Two-sample $Z$-test (5.1)
- **(b)** Two guardrail configurations; benchmark scores from $n_1 = 9$ and $n_2 = 11$ independent eval runs; $s_1 \approx s_2$, variances unknown.

# Quick Practice: Which Two-Sample Tool?

*AI safety testing scenarios*

- **(a)** Two AV software builds; disengagement-free distance per intervention, $\sigma_1, \sigma_2$ known from years of fleet logs; large independent samples.
  $\Rightarrow$ Two-sample $Z$-test (5.1)
- **(b)** Two guardrail configurations; benchmark scores from $n_1 = 9$ and $n_2 = 11$ independent eval runs; $s_1 \approx s_2$, variances unknown.
  $\Rightarrow$ Pooled $t$-test (5.4), $\nu = 18$

# Quick Practice: Which Two-Sample Tool?

*AI safety testing scenarios*

- **(a)** Two AV software builds; disengagement-free distance per intervention, $\sigma_1, \sigma_2$ known from years of fleet logs; large independent samples.
  $\Rightarrow$ Two-sample $Z$-test (5.1)
- **(b)** Two guardrail configurations; benchmark scores from $n_1 = 9$ and $n_2 = 11$ independent eval runs; $s_1 \approx s_2$, variances unknown.
  $\Rightarrow$ Pooled $t$-test (5.4), $\nu = 18$
- **(c)** The *same* 20 red-team prompt suites scored before and after fine-tuning the model.

# Quick Practice: Which Two-Sample Tool?

*AI safety testing scenarios*

- **(a)** Two AV software builds; disengagement-free distance per intervention, $\sigma_1, \sigma_2$ known from years of fleet logs; large independent samples.
  $\Rightarrow$ Two-sample $Z$-test (5.1)
- **(b)** Two guardrail configurations; benchmark scores from $n_1 = 9$ and $n_2 = 11$ independent eval runs; $s_1 \approx s_2$, variances unknown.
  $\Rightarrow$ Pooled $t$-test (5.4), $\nu = 18$
- **(c)** The *same* 20 red-team prompt suites scored before and after fine-tuning the model.
  $\Rightarrow$ Not independent samples — paired $t$-test, next lecture

# Where We Are: The Growing Test Menu

| Parameter | Conditions | Test Statistic | Distribution | Book Eq. |
|---|---|---|---|---|
| *Single sample (Ch. 3–4):* | | | | |
| $\mu$ | $\sigma$ known | $Z_0$ | $N(0,1)$ | (4.3) |
| $\mu$ | $\sigma$ unknown | $T_0$ | $t_{n-1}$ | (4.6) |
| $\sigma^2$ | Normal pop. | $\chi^2_0$ | $\chi^2_{n-1}$ | (4.7) |
| $p$ | Large $n$ | $Z_0$ | $N(0,1)$ | (4.8) |
| *Two samples (Ch. 5):* | | | | |
| $\mu_1 - \mu_2$ | $\sigma_1, \sigma_2$ known | $Z_0$ | $N(0,1)$ | (5.1) |
| $\mu_1 - \mu_2$ | $\sigma_1^2 = \sigma_2^2$ unknown | $T_0$ (pooled) | $t_{n_1+n_2-2}$ | (5.4) |

# Where We Are: The Growing Test Menu

| Parameter | Conditions | Test Statistic | Distribution | Book Eq. |
|---|---|---|---|---|
| *Single sample (Ch. 3–4):* | | | | |
| $\mu$ | $\sigma$ known | $Z_0$ | $N(0,1)$ | (4.3) |
| $\mu$ | $\sigma$ unknown | $T_0$ | $t_{n-1}$ | (4.6) |
| $\sigma^2$ | Normal pop. | $\chi^2_0$ | $\chi^2_{n-1}$ | (4.7) |
| $p$ | Large $n$ | $Z_0$ | $N(0,1)$ | (4.8) |
| *Two samples (Ch. 5):* | | | | |
| $\mu_1 - \mu_2$ | $\sigma_1, \sigma_2$ known | $Z_0$ | $N(0,1)$ | (5.1) |
| $\mu_1 - \mu_2$ | $\sigma_1^2 = \sigma_2^2$ unknown | $T_0$ (pooled) | $t_{n_1+n_2-2}$ | (5.4) |
| $\mu_1 - \mu_2$ | $\sigma_1^2 \neq \sigma_2^2$ unknown | $T_0$ (Welch) | *Lecture 12* | (5.6) |
| $\mu_D$ (paired) | Paired data | $T_0$ | *Lecture 12* | (5.8) |

## Common Mistakes to Avoid

- **Using pooled $t$ when variances are clearly unequal.**
  Check $s_1$ and $s_2$. If one is more than twice the other, be cautious — Welch (next lecture).

## Common Mistakes to Avoid

- **Using pooled $t$ when variances are clearly unequal.**
  Check $s_1$ and $s_2$. If one is more than twice the other, be cautious — Welch (next lecture).
- **Wrong degrees of freedom.**
  Pooled $t$ (5.4): $\nu = n_1 + n_2 - 2$, *not* $n_1 - 1$ or $n_2 - 1$.

## Common Mistakes to Avoid

- **Using pooled $t$ when variances are clearly unequal.**
  Check $s_1$ and $s_2$. If one is more than twice the other, be cautious — Welch (next lecture).
- **Wrong degrees of freedom.**
  Pooled $t$ (5.4): $\nu = n_1 + n_2 - 2$, *not* $n_1 - 1$ or $n_2 - 1$.
- **Forgetting $\Delta_0$ in the numerator.**
  Usually $\Delta_0 = 0$, but read carefully. Sometimes $H_0 : \mu_1 - \mu_2 = 5$.

## Common Mistakes to Avoid

- **Using pooled $t$ when variances are clearly unequal.**
  Check $s_1$ and $s_2$. If one is more than twice the other, be cautious — Welch (next lecture).
- **Wrong degrees of freedom.**
  Pooled $t$ (5.4): $\nu = n_1 + n_2 - 2$, *not* $n_1 - 1$ or $n_2 - 1$.
- **Forgetting $\Delta_0$ in the numerator.**
  Usually $\Delta_0 = 0$, but read carefully. Sometimes $H_0 : \mu_1 - \mu_2 = 5$.
- **Using a two-sample test for paired data.**
  If observations are naturally paired, use the paired $t$-test (5.8) (Lecture 12).

## Lecture 11 Summary

- Chapter 5 extends inference from one sample to **two independent samples**.

## Lecture 11 Summary

- Chapter 5 extends inference from one sample to **two independent samples**.
- $\sigma_1, \sigma_2$ **known**: two-sample $Z$-test (5.1) and $Z$-CI (5.2).

# Lecture 11 Summary

- Chapter 5 extends inference from one sample to **two independent samples**.
- $\sigma_1, \sigma_2$ **known**: two-sample $Z$-test (5.1) and $Z$-CI (5.2).
- Variances **unknown but equal**: pool via (5.3), then the pooled $t$-test (5.4) and $t$-CI (5.5), with $\nu = n_1 + n_2 - 2$.

# Lecture 11 Summary

- Chapter 5 extends inference from one sample to **two independent samples**.
- $\sigma_1, \sigma_2$ **known**: two-sample $Z$-test (5.1) and $Z$-CI (5.2).
- Variances **unknown but equal**: pool via (5.3), then the pooled $t$-test (5.4) and $t$-CI (5.5), with $\nu = n_1 + n_2 - 2$.
- The seven-step procedure (Procedure 4.2) is **unchanged**. Only the test statistic and its degrees of freedom change.

## Lecture 11 Summary

- Chapter 5 extends inference from one sample to **two independent samples**.
- $\sigma_1, \sigma_2$ **known**: two-sample $Z$-test (5.1) and $Z$-CI (5.2).
- Variances **unknown but equal**: pool via (5.3), then the pooled $t$-test (5.4) and $t$-CI (5.5), with $\nu = n_1 + n_2 - 2$.
- The seven-step procedure (Procedure 4.2) is **unchanged**. Only the test statistic and its degrees of freedom change.
- CI–test duality still holds: $\Delta_0$ outside the CI $\Leftrightarrow$ reject $H_0$.

## Lecture 11 Summary

- Chapter 5 extends inference from one sample to **two independent samples**.
- $\sigma_1, \sigma_2$ **known**: two-sample $Z$-test (5.1) and $Z$-CI (5.2).
- Variances **unknown but equal**: pool via (5.3), then the pooled $t$-test (5.4) and $t$-CI (5.5), with $\nu = n_1 + n_2 - 2$.
- The seven-step procedure (Procedure 4.2) is **unchanged**. Only the test statistic and its degrees of freedom change.
- CI–test duality still holds: $\Delta_0$ outside the CI $\Leftrightarrow$ reject $H_0$.
- **Next lecture:** Welch's $t$ (5.6), paired $t$ (5.8), the $F$-test (5.13), and two proportions (5.10).