ISE 315: Engineering Statistics

Lecture 12: Statistical Inference for Two Samples (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 12

Two Samples: Welch's t, Paired t, F-Test, and Two Proportions
(Chapter 5)

Lecture 12 Outline

- Quick review of Lecture 11 (two-sample Z and pooled t)

Lecture 12 Outline

- Quick review of Lecture 11 (two-sample Z and pooled t)
- Welch's t-test: unequal variances (§5.4)

Lecture 12 Outline

- Quick review of Lecture 11 (two-sample Z and pooled t)
- Welch's t-test: unequal variances (§5.4)
- The paired t-test (§5.5)

Lecture 12 Outline

- Quick review of Lecture 11 (two-sample Z and pooled t)
- Welch's t-test: unequal variances (§5.4)
- The paired t-test (§5.5)
- Inference on two population proportions (§5.6)

Lecture 12 Outline

- Quick review of Lecture 11 (two-sample Z and pooled t)
- Welch's t-test: unequal variances (§5.4)
- The paired t-test (§5.5)
- Inference on two population proportions (§5.6)
- F-test for the ratio of two variances (§5.7)

Lecture 12 Outline

- Quick review of Lecture 11 (two-sample Z and pooled t)
- Welch's t-test: unequal variances (§5.4)
- The paired t-test (§5.5)
- Inference on two population proportions (§5.6)
- F-test for the ratio of two variances (§5.7)
- Choosing among the two-sample tools (Procedure 5.2)

Review: What We Covered in Lecture 11

Two-sample Z**-test** (σ_1, σ_2 known) — (5.1), CI (5.2):

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

Review: What We Covered in Lecture 11

Two-sample Z**-test** (σ_1, σ_2 known) — (5.1), CI (5.2):

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

Pooled t**-test** ($\sigma_1^2 = \sigma_2^2$ assumed, both unknown) — (5.3), (5.4), CI (5.5):

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Review: What We Covered in Lecture 11

Two-sample Z-test (σ_1, σ_2 known) — (5.1), CI (5.2):

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

Pooled t-test ($\sigma_1^2 = \sigma_2^2$ assumed, both unknown) — (5.3), (5.4), CI (5.5):

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}, \quad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Today: What if $\sigma_1^2 \neq \sigma_2^2$? What if the samples are paired? What about proportions and variances?

When Variances Are Unequal

Welch's t-Test

If $\sigma_1^2 \neq \sigma_2^2$ (or we cannot assume they are equal), **do NOT pool the variances!**

Textbook: Chapter 5, §5.4 ((5.6) onward)

When Variances Are Unequal

Welch's t-Test

If $\sigma_1^2 \neq \sigma_2^2$ (or we cannot assume they are equal), **do NOT pool the variances!**

Welch's t-test: Keep the sample variances separate.

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{5.6}$$

Textbook: Chapter 5, §5.4 ((5.6) onward)

When Variances Are Unequal

Welch's t-Test

If $\sigma_1^2 \neq \sigma_2^2$ (or we cannot assume they are equal), **do NOT pool the variances!**

Welch's t-test: Keep the sample variances separate.

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{5.6}$$

The distribution of T_0 is approximately t_ν, where the degrees of freedom ν comes from the **Welch–Satterthwaite** formula (next slide).

Textbook: Chapter 5, §5.4 ((5.6) onward)

Welch–Satterthwaite Degrees of Freedom

See also Figure 5.3 in the textbook

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{5.7}$$

Welch–Satterthwaite Degrees of Freedom

See also Figure 5.3 in the textbook

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{5.7}$$

Always round down to the nearest integer (the conservative choice).

Welch–Satterthwaite Degrees of Freedom

See also Figure 5.3 in the textbook

$$\nu = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{5.7}$$

Always round down to the nearest integer (the conservative choice).

- If $s_1^2 \approx s_2^2$ and $n_1 \approx n_2$, then $\nu \approx n_1 + n_2 - 2$ (close to pooled).

Welch–Satterthwaite Degrees of Freedom

See also Figure 5.3 in the textbook

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{5.7}$$

Always round down to the nearest integer (the conservative choice).

- If $s_1^2 \approx s_2^2$ and $n_1 \approx n_2$, then $\nu \approx n_1 + n_2 - 2$ (close to pooled).
- If the variances are very different, ν can be much smaller than $n_1 + n_2 - 2$.

Welch–Satterthwaite Degrees of Freedom

See also Figure 5.3 in the textbook

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{5.7}$$

Always round down to the nearest integer (the conservative choice).

- If $s_1^2 \approx s_2^2$ and $n_1 \approx n_2$, then $\nu \approx n_1 + n_2 - 2$ (close to pooled).
- If the variances are very different, ν can be much smaller than $n_1 + n_2 - 2$.
- Smaller df $\Rightarrow$ wider confidence intervals and harder to reject H_0.

Confidence Interval: Welch's t $(\sigma_1^2 \neq \sigma_2^2)$

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

$$(\bar{x}_1 - \bar{x}_2) \;\pm\; t_{\alpha/2,\,\nu} \;\cdot\; \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where ν is from the Welch–Satterthwaite formula (5.7).

Confidence Interval: Welch's t ($\sigma_1^2 \neq \sigma_2^2$)

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

$$(\bar{x}_1 - \bar{x}_2) \;\pm\; t_{\alpha/2,\nu} \;\cdot\; \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where ν is from the Welch–Satterthwaite formula (5.7).

Comparison of standard errors:

Method	**Standard Error**	**df**
Pooled t (5.4)	$s_p\sqrt{1/n_1 + 1/n_2}$	$n_1 + n_2 - 2$
Welch's t (5.6)	$\sqrt{s_1^2/n_1 + s_2^2/n_2}$	ν from (5.7)

Example: Comparing Two Model Versions on Eval Suites

Example 5.3 in the textbook

An AI safety team scores two model versions on independent evaluation suites (safety score, 0–100).

- Model A: $n_1 = 10$, $\bar{x}_1 = 78.4$, $s_1 = 3.1$
- Model B: $n_2 = 16$, $\bar{x}_2 = 74.2$, $s_2 = 7.9$

Do the mean safety scores differ at $\alpha = 0.05$?

Example: Comparing Two Model Versions on Eval Suites

Example 5.3 in the textbook

An AI safety team scores two model versions on independent evaluation suites (safety score, 0–100).

- Model A: $n_1 = 10$, $\bar{x}_1 = 78.4$, $s_1 = 3.1$
- Model B: $n_2 = 16$, $\bar{x}_2 = 74.2$, $s_2 = 7.9$

Do the mean safety scores differ at $\alpha = 0.05$?

Check the s ratio first: $s_2/s_1 = 7.9/3.1 = 2.548 > 2$. Do **not** assume equal variances.

Example: Comparing Two Model Versions on Eval Suites

Example 5.3 in the textbook

An AI safety team scores two model versions on independent evaluation suites (safety score, 0–100).

- Model A: $n_1 = 10$, $\bar{x}_1 = 78.4$, $s_1 = 3.1$
- Model B: $n_2 = 16$, $\bar{x}_2 = 74.2$, $s_2 = 7.9$

Do the mean safety scores differ at $\alpha = 0.05$?

Check the s ratio first: $s_2/s_1 = 7.9/3.1 = 2.548 > 2$. Do **not** assume equal variances.

Steps 1–3: Parameter: $\mu_1 - \mu_2$. Variances unknown and unequal $\Rightarrow$ Welch's t-test (5.6).

Example: Comparing Two Model Versions on Eval Suites

Example 5.3 in the textbook

An AI safety team scores two model versions on independent evaluation suites (safety score, 0–100).

- Model A: $n_1 = 10$, $\bar{x}_1 = 78.4$, $s_1 = 3.1$
- Model B: $n_2 = 16$, $\bar{x}_2 = 74.2$, $s_2 = 7.9$

Do the mean safety scores differ at $\alpha = 0.05$?

Check the s ratio first: $s_2/s_1 = 7.9/3.1 = 2.548 > 2$. Do **not** assume equal variances.

Steps 1–3: Parameter: $\mu_1 - \mu_2$. Variances unknown and unequal $\Rightarrow$ Welch's t-test (5.6).

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

Example: Model Versions (Solution)

Step 6a: Standard error: $\sqrt{\frac{3.1^2}{10} + \frac{7.9^2}{16}} = \sqrt{0.9610 + 3.9006} = \sqrt{4.8616} =$ 2.205

Example: Model Versions (Solution)

Step 6a: Standard error: $\sqrt{\frac{3.1^2}{10} + \frac{7.9^2}{16}} = \sqrt{0.9610 + 3.9006} = \sqrt{4.8616} = 2.205$

Step 6b: $T_0 = \frac{(78.4 - 74.2) - 0}{2.205} = \frac{4.2}{2.205} = 1.905$

Example: Model Versions (Solution)

Step 6a: Standard error: $\sqrt{\frac{3.1^2}{10}+\frac{7.9^2}{16}} = \sqrt{0.9610+3.9006} = \sqrt{4.8616} = 2.205$

Step 6b: $T_0 = \frac{(78.4-74.2)-0}{2.205} = \frac{4.2}{2.205} = 1.905$

Step 6c: Degrees of freedom (5.7):

$$\nu = \frac{(0.9610+3.9006)^2}{\frac{0.9610^2}{9}+\frac{3.9006^2}{15}} = \frac{23.64}{0.1026+1.0143} = \frac{23.64}{1.1169} = 21.16 \Rightarrow \nu = 21$$

Example: Model Versions (Solution)

Step 6a: Standard error: $\sqrt{\dfrac{3.1^2}{10} + \dfrac{7.9^2}{16}} = \sqrt{0.9610 + 3.9006} = \sqrt{4.8616} = 2.205$

Step 6b: $T_0 = \dfrac{(78.4 - 74.2) - 0}{2.205} = \dfrac{4.2}{2.205} = 1.905$

Step 6c: Degrees of freedom (5.7):

$$\nu = \frac{(0.9610 + 3.9006)^2}{\dfrac{0.9610^2}{9} + \dfrac{3.9006^2}{15}} = \frac{23.64}{0.1026 + 1.0143} = \frac{23.64}{1.1169} = 21.16 \;\Rightarrow\; \nu = 21$$

Steps 4–5, 7: $t_{0.025,21} = 2.080$. $|T_0| = 1.905 < 2.080 \Rightarrow$ **fail to reject H_0**.
P-value bounds: $t_{0.05,21} = 1.721 < 1.905 < 2.080$, so $0.05 <$ P-value < 0.10.
95% CI: $4.2 \pm 2.080\,(2.205) = (-0.386,\ 8.786)$ — contains 0, consistent.

The Paired t-Test (When Observations Come in Pairs)

Sometimes the two samples are **not independent**. Each observation in sample 1 is naturally paired with an observation in sample 2.

Textbook: Chapter 5, §5.5 ((5.8) onward); see Figure 5.4

The Paired t-Test (When Observations Come in Pairs)

Sometimes the two samples are **not independent**. Each observation in sample 1 is naturally paired with an observation in sample 2.

Common examples:

- Before/after measurements on the **same** unit (truck, model, user)
- Two methods applied to the **same** specimen
- Matched subjects (same route, same prompt suite, same customer cohort)

Textbook: Chapter 5, §5.5 ((5.8) onward); see Figure 5.4

The Paired t-Test (When Observations Come in Pairs)

Sometimes the two samples are **not independent**. Each observation in sample 1 is naturally paired with an observation in sample 2.

Common examples:

- Before/after measurements on the **same** unit (truck, model, user)
- Two methods applied to the **same** specimen
- Matched subjects (same route, same prompt suite, same customer cohort)

Key idea: Do not treat them as two samples — analyze the **differences** $D_j = X_{1j} - X_{2j}$ as a *single sample*.

Textbook: Chapter 5, §5.5 ((5.8) onward); see Figure 5.4

The Paired t-Test (When Observations Come in Pairs)

Sometimes the two samples are **not independent**. Each observation in sample 1 is naturally paired with an observation in sample 2.

Common examples:

- Before/after measurements on the **same** unit (truck, model, user)
- Two methods applied to the **same** specimen
- Matched subjects (same route, same prompt suite, same customer cohort)

Key idea: Do not treat them as two samples — analyze the **differences** $D_j = X_{1j} - X_{2j}$ as a *single sample*.

This removes unit-to-unit variability, making the comparison **more precise**.

Textbook: Chapter 5, §5.5 ((5.8) onward); see Figure 5.4

Paired t-Test: Formulas

Given n paired observations, compute the differences $d_j = x_{1j} - x_{2j}$, then:

$$\bar{D} = \frac{1}{n}\sum_{j=1}^{n} d_j, \qquad S_D = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(d_j - \bar{D})^2}$$

Paired t-Test: Formulas

Given n paired observations, compute the differences $d_j = x_{1j} - x_{2j}$, then:

$$\bar{D} = \frac{1}{n}\sum_{j=1}^{n} d_j, \qquad S_D = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(d_j - \bar{D})^2}$$

Hypotheses: $H_0 : \mu_D = \Delta_0$ vs $H_1 : \mu_D \neq \Delta_0$ (often $\Delta_0 = 0$)

Paired t-Test: Formulas

Given n paired observations, compute the differences $d_j = x_{1j} - x_{2j}$, then:

$$\bar{D} = \frac{1}{n}\sum_{j=1}^{n} d_j, \qquad S_D = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(d_j - \bar{D})^2}$$

Hypotheses: $H_0 : \mu_D = \Delta_0$ vs $H_1 : \mu_D \neq \Delta_0$ (often $\Delta_0 = 0$)

Test statistic:

$$T_0 = \frac{\bar{D} - \Delta_0}{S_D/\sqrt{n}} \tag{5.8}$$

Under H_0, $T_0 \sim t_{n-1}$ $\quad$ (n = number of **pairs**).

Paired t-Test: Formulas

Given n paired observations, compute the differences $d_j = x_{1j} - x_{2j}$, then:

$$\bar{D} = \frac{1}{n}\sum_{j=1}^{n} d_j, \qquad S_D = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(d_j - \bar{D})^2}$$

Hypotheses: $H_0 : \mu_D = \Delta_0$ vs $H_1 : \mu_D \neq \Delta_0$ (often $\Delta_0 = 0$)

Test statistic:

$$T_0 = \frac{\bar{D} - \Delta_0}{S_D/\sqrt{n}} \tag{5.8}$$

Under H_0, $T_0 \sim t_{n-1}$ $\quad$ (n = number of **pairs**).

CI for μ_D:

$$\bar{d} \;\pm\; t_{\alpha/2,\, n-1}\, \frac{s_D}{\sqrt{n}} \tag{5.9}$$

Example: Truck Fuel Efficiency With Route-Optimization Software

Example 5.4 in the textbook

Eight delivery trucks run their weekly cycle without, then with, route-optimization software. Fuel efficiency in km/L:

Truck	1	2	3	4	5	6	7	8
Without (x_{1j})	2.9	3.4	2.7	3.1	3.6	2.8	3.3	3.0
With (x_{2j})	3.5	3.7	3.5	3.6	4.0	3.5	3.5	3.5

Example: Truck Fuel Efficiency With Route-Optimization Software

Example 5.4 in the textbook

Eight delivery trucks run their weekly cycle without, then with, route-optimization software. Fuel efficiency in km/L:

Truck	1	2	3	4	5	6	7	8
Without (x_{1j})	2.9	3.4	2.7	3.1	3.6	2.8	3.3	3.0
With (x_{2j})	3.5	3.7	3.5	3.6	4.0	3.5	3.5	3.5
$d_j = x_{2j} - x_{1j}$	0.6	0.3	0.8	0.5	0.4	0.7	0.2	0.5

Example: Truck Fuel Efficiency With Route-Optimization Software

Example 5.4 in the textbook

Eight delivery trucks run their weekly cycle without, then with, route-optimization software. Fuel efficiency in km/L:

Truck	1	2	3	4	5	6	7	8
Without (x_{1j})	2.9	3.4	2.7	3.1	3.6	2.8	3.3	3.0
With (x_{2j})	3.5	3.7	3.5	3.6	4.0	3.5	3.5	3.5
$d_j = x_{2j} - x_{1j}$	0.6	0.3	0.8	0.5	0.4	0.7	0.2	0.5

Does the software improve fuel efficiency? Use $\alpha = 0.05$.

Note: Paired data (same truck, two conditions) $\Rightarrow$ paired t-test (5.8).
Let d_j = With − Without. $H_0 : \mu_D = 0$ vs $H_1 : \mu_D > 0$ (right-tailed).

Example: Truck Fuel Efficiency (Solution)

Step 6a: $\bar{d} = \dfrac{0.6 + 0.3 + 0.8 + 0.5 + 0.4 + 0.7 + 0.2 + 0.5}{8} = \dfrac{4.0}{8} = 0.5$

Example: Truck Fuel Efficiency (Solution)

Step 6a: $\bar{d} = \dfrac{0.6 + 0.3 + 0.8 + 0.5 + 0.4 + 0.7 + 0.2 + 0.5}{8} = \dfrac{4.0}{8} = 0.5$

Step 6b: $\sum (d_j - \bar{d})^2 = 0.01 + 0.04 + 0.09 + 0 + 0.01 + 0.04 + 0.09 + 0 = 0.28$

$s_D = \sqrt{0.28/7} = \sqrt{0.04} = 0.2$

Example: Truck Fuel Efficiency (Solution)

Step 6a: $\bar{d} = \dfrac{0.6 + 0.3 + 0.8 + 0.5 + 0.4 + 0.7 + 0.2 + 0.5}{8} = \dfrac{4.0}{8} = 0.5$

Step 6b: $\sum (d_j - \bar{d})^2 = 0.01 + 0.04 + 0.09 + 0 + 0.01 + 0.04 + 0.09 + 0 = 0.28$

$s_D = \sqrt{0.28/7} = \sqrt{0.04} = 0.2$

Step 6c: Test statistic (5.8): $T_0 = \dfrac{0.5 - 0}{0.2/\sqrt{8}} = \dfrac{0.5}{0.07071} = 7.071$

Example: Truck Fuel Efficiency (Solution)

Step 6a: $\bar{d} = \dfrac{0.6 + 0.3 + 0.8 + 0.5 + 0.4 + 0.7 + 0.2 + 0.5}{8} = \dfrac{4.0}{8} = 0.5$

Step 6b: $\sum(d_j - \bar{d})^2 = 0.01 + 0.04 + 0.09 + 0 + 0.01 + 0.04 + 0.09 + 0 = 0.28$
$s_D = \sqrt{0.28/7} = \sqrt{0.04} = 0.2$

Step 6c: Test statistic (5.8): $T_0 = \dfrac{0.5 - 0}{0.2/\sqrt{8}} = \dfrac{0.5}{0.07071} = 7.071$

Steps 4–5, 7: $\nu = 7$; $t_{0.05,7} = 1.895$. $T_0 = 7.071 \gg 1.895$; from the table $t_{0.005,7} = 3.499$, so P-value < 0.005.
Reject H_0. Strong evidence the software improves fuel efficiency.

Example: Truck Fuel Efficiency (Solution)

Step 6a: $\bar{d} = \dfrac{0.6 + 0.3 + 0.8 + 0.5 + 0.4 + 0.7 + 0.2 + 0.5}{8} = \dfrac{4.0}{8} = 0.5$

Step 6b: $\sum(d_j - \bar{d})^2 = 0.01 + 0.04 + 0.09 + 0 + 0.01 + 0.04 + 0.09 + 0 = 0.28$
$s_D = \sqrt{0.28/7} = \sqrt{0.04} = 0.2$

Step 6c: Test statistic (5.8): $T_0 = \dfrac{0.5 - 0}{0.2/\sqrt{8}} = \dfrac{0.5}{0.07071} = 7.071$

Steps 4–5, 7: $\nu = 7$; $t_{0.05,7} = 1.895$. $T_0 = 7.071 \gg 1.895$; from the table $t_{0.005,7} = 3.499$, so P-value < 0.005.
Reject H_0. Strong evidence the software improves fuel efficiency.

95% CI for μ_D (5.9): $0.5 \pm 2.365\,(0.07071) = 0.5 \pm 0.167 = (0.333,\ 0.667)$ km/L.

Paired vs. Independent: How to Decide

See Table 5.4 in the textbook

	Two-Sample	**Paired**
Design	Two separate groups	Same units, two conditions
Sample sizes	Can differ ($n_1 \neq n_2$)	Always equal (n pairs)
Parameter	$\mu_1 - \mu_2$	μ_D
df	$n_1 + n_2 - 2$ (pooled) or Welch (5.7)	$n - 1$
Advantage	Simpler design	No unit-to-unit variability

Paired vs. Independent: How to Decide

See Table 5.4 in the textbook

	Two-Sample	**Paired**
Design	Two separate groups	Same units, two conditions
Sample sizes	Can differ ($n_1 \neq n_2$)	Always equal (n pairs)
Parameter	$\mu_1 - \mu_2$	μ_D
df	$n_1 + n_2 - 2$ (pooled) or Welch (5.7)	$n - 1$
Advantage	Simpler design	No unit-to-unit variability

Inference on Two Population Proportions

Two independent random samples:

- Sample 1: n_1 trials, x_1 successes, $\hat{p}_1 = x_1/n_1$
- Sample 2: n_2 trials, x_2 successes, $\hat{p}_2 = x_2/n_2$

Inference on Two Population Proportions

Two independent random samples:

- Sample 1: n_1 trials, x_1 successes, $\hat{p}_1 = x_1/n_1$
- Sample 2: n_2 trials, x_2 successes, $\hat{p}_2 = x_2/n_2$

Hypotheses:

$$H_0 : p_1 = p_2 \quad \text{vs} \quad H_1 : p_1 \neq p_2 \qquad (\text{equivalently } H_0 : p_1 - p_2 = 0)$$

Inference on Two Population Proportions

Two independent random samples:

- Sample 1: n_1 trials, x_1 successes, $\hat{p}_1 = x_1/n_1$
- Sample 2: n_2 trials, x_2 successes, $\hat{p}_2 = x_2/n_2$

Hypotheses:

$$H_0 : p_1 = p_2 \quad \text{vs} \quad H_1 : p_1 \neq p_2 \qquad \text{(equivalently } H_0 : p_1 - p_2 = 0\text{)}$$

Under H_0: Since $p_1 = p_2 = p$ (a common value), we estimate p by **pooling**:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Test Statistic for Two Proportions

Test statistic:

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{5.10}$$

where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$ is the pooled proportion.

Test Statistic for Two Proportions

Test statistic:

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{5.10}$$

where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$ is the pooled proportion.

Under H_0, $Z_0 \sim N(0,1)$ approximately, for large samples ($n_i\hat{p}_i \geq 5$ and $n_i(1-\hat{p}_i) \geq 5$).

Test Statistic for Two Proportions

Test statistic:

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{5.10}$$

where $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ is the pooled proportion.

Under H_0, $Z_0 \sim N(0,1)$ approximately, for large samples ($n_i\hat{p}_i \geq 5$ and $n_i(1-\hat{p}_i) \geq 5$).

Decision rules are identical to any Z-test:
Two-tailed: reject if $|Z_0| > z_{\alpha/2}$. One-tailed: reject if $Z_0 > z_\alpha$ (or $< -z_\alpha$).

Confidence Interval for $p_1 - p_2$

A $100(1-\alpha)\%$ CI for $p_1 - p_2$:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \tag{5.11}$$

Confidence Interval for $p_1 - p_2$

A $100(1-\alpha)\%$ CI for $p_1 - p_2$:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \tag{5.11}$$

Important: The CI uses the **individual** sample proportions $\hat{p}_1, \hat{p}_2$ in the standard error, *not* the pooled $\hat{p}$. The pooled $\hat{p}$ belongs to the hypothesis test (5.10) only.

Confidence Interval for $p_1 - p_2$

A $100(1-\alpha)\%$ CI for $p_1 - p_2$:

$$(\hat{p}_1 - \hat{p}_2) \ \pm \ z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \tag{5.11}$$

Important: The CI uses the **individual** sample proportions $\hat{p}_1, \hat{p}_2$ in the standard error, *not* the pooled $\hat{p}$. The pooled $\hat{p}$ belongs to the hypothesis test (5.10) only.

Duality check: If this interval contains 0, the two-tailed test at level α fails to reject (approximately — the two standard errors differ slightly).

Example: Guardrail False-Positive Rates, v1 vs v2

Example 5.5 in the textbook

An AI safety team compares the false-positive rates of two guardrail versions on benign prompts.

- Version 1: $n_1 = 400$ prompts, $x_1 = 52$ falsely blocked, $\hat{p}_1 = 0.13$
- Version 2: $n_2 = 500$ prompts, $x_2 = 45$ falsely blocked, $\hat{p}_2 = 0.09$

Do the false-positive rates differ? Use $\alpha = 0.05$.

Example: Guardrail False-Positive Rates, v1 vs v2

Example 5.5 in the textbook

An AI safety team compares the false-positive rates of two guardrail versions on benign prompts.

- Version 1: $n_1 = 400$ prompts, $x_1 = 52$ falsely blocked, $\hat{p}_1 = 0.13$
- Version 2: $n_2 = 500$ prompts, $x_2 = 45$ falsely blocked, $\hat{p}_2 = 0.09$

Do the false-positive rates differ? Use $\alpha = 0.05$.

Steps 1–3: $H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$, $\alpha = 0.05$. Two-proportion Z-test (5.10). Counts $52, 348, 45, 455$ all ≥ 5: large-sample conditions hold.

Example: Guardrail False-Positive Rates, v1 vs v2

Example 5.5 in the textbook

An AI safety team compares the false-positive rates of two guardrail versions on benign prompts.

- Version 1: $n_1 = 400$ prompts, $x_1 = 52$ falsely blocked, $\hat{p}_1 = 0.13$
- Version 2: $n_2 = 500$ prompts, $x_2 = 45$ falsely blocked, $\hat{p}_2 = 0.09$

Do the false-positive rates differ? Use $\alpha = 0.05$.

Steps 1–3: $H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$, $\alpha = 0.05$. Two-proportion Z-test (5.10). Counts $52, 348, 45, 455$ all ≥ 5: large-sample conditions hold.

Step 6a: Pooled proportion: $\hat{p} = \dfrac{52 + 45}{400 + 500} = \dfrac{97}{900} = 0.1078$

Example: Guardrail False-Positive Rates, v1 vs v2

Example 5.5 in the textbook

An AI safety team compares the false-positive rates of two guardrail versions on benign prompts.

- Version 1: $n_1 = 400$ prompts, $x_1 = 52$ falsely blocked, $\hat{p}_1 = 0.13$
- Version 2: $n_2 = 500$ prompts, $x_2 = 45$ falsely blocked, $\hat{p}_2 = 0.09$

Do the false-positive rates differ? Use $\alpha = 0.05$.

Steps 1–3: $H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$, $\alpha = 0.05$. Two-proportion Z-test (5.10). Counts $52, 348, 45, 455$ all ≥ 5: large-sample conditions hold.

Step 6a: Pooled proportion: $\hat{p} = \dfrac{52 + 45}{400 + 500} = \dfrac{97}{900} = 0.1078$

Step 6b: $Z_0 = \dfrac{0.13 - 0.09}{\sqrt{0.1078\,(0.8922)\left(\frac{1}{400} + \frac{1}{500}\right)}} = \dfrac{0.04}{\sqrt{0.0004327}} = \dfrac{0.04}{0.02080} =$

Example: Guardrail False Positives (Conclusion)

Steps 4–5: Two-tailed, $z_{0.025} = 1.96$. Rejection region: $|Z_0| > 1.96$.

Example: Guardrail False Positives (Conclusion)

Steps 4–5: Two-tailed, $z_{0.025} = 1.96$. Rejection region: $|Z_0| > 1.96$.

Decision: $|Z_0| = 1.923 < 1.96$. Not in the rejection region.
P-value $= 2[1 - \Phi(1.92)] = 2(1 - 0.9726) = 2(0.0274) = 0.0548$

Example: Guardrail False Positives (Conclusion)

Steps 4–5: Two-tailed, $z_{0.025} = 1.96$. Rejection region: $|Z_0| > 1.96$.

Decision: $|Z_0| = 1.923 < 1.96$. Not in the rejection region.
P-value $= 2[1 - \Phi(1.92)] = 2(1 - 0.9726) = 2(0.0274) = 0.0548$

Step 7: **Fail to reject H_0** — barely. At $\alpha = 0.05$ there is not quite sufficient evidence that the false-positive rates differ. Worth more testing before shipping v2.

Example: Guardrail False Positives (Conclusion)

Steps 4–5: Two-tailed, $z_{0.025} = 1.96$. Rejection region: $|Z_0| > 1.96$.

Decision: $|Z_0| = 1.923 < 1.96$. Not in the rejection region.
P-value $= 2[1 - \Phi(1.92)] = 2(1 - 0.9726) = 2(0.0274) = 0.0548$

Step 7: **Fail to reject** H_0 — barely. At $\alpha = 0.05$ there is not quite sufficient evidence that the false-positive rates differ. Worth more testing before shipping v2.

95% CI for $p_1 - p_2$ (5.11):

$$(0.13 - 0.09) \pm 1.96\sqrt{\frac{0.13(0.87)}{400} + \frac{0.09(0.91)}{500}}$$
$$= 0.04 \pm 1.96\,(0.02113) = 0.04 \pm 0.0414 = (-0.0014,\ 0.0814)$$

Contains 0 (just), consistent with failing to reject.

Comparing Two Variances: The F-Test

Sometimes we compare the **variability** of two populations, not their means.

- Is one process more consistent than another?
- Should we use the pooled t-test or Welch's t-test?

Textbook: Chapter 5, §5.7 ((5.12) onward)

Comparing Two Variances: The F-Test

Sometimes we compare the **variability** of two populations, not their means.

- Is one process more consistent than another?
- Should we use the pooled t-test or Welch's t-test?

The F distribution: For independent normal samples,

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

Textbook: Chapter 5, §5.7 ((5.12) onward)

Comparing Two Variances: The F-Test

Sometimes we compare the **variability** of two populations, not their means.

- Is one process more consistent than another?
- Should we use the pooled t-test or Welch's t-test?

The F distribution: For independent normal samples,

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

The F distribution has **two** degree-of-freedom parameters: numerator df $u = n_1 - 1$ and denominator df $v = n_2 - 1$. See Figure 5.5 in the textbook.

Textbook: Chapter 5, §5.7 ((5.12) onward)

F-Test for σ_1^2/σ_2^2

Procedure 5.1

Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2 \qquad (\text{equivalently } H_0 : \sigma_1^2/\sigma_2^2 = 1)$$

F-Test for σ_1^2/σ_2^2

Procedure 5.1

Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2 \qquad (\text{equivalently } H_0 : \sigma_1^2/\sigma_2^2 = 1)$$

Test statistic:

$$F_0 = \frac{S_1^2}{S_2^2} \tag{5.13}$$

Under H_0, $F_0 \sim F_{n_1-1,\, n_2-1}$.

F-Test for σ_1^2/σ_2^2

Procedure 5.1

Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2 \qquad (\text{equivalently } H_0 : \sigma_1^2/\sigma_2^2 = 1)$$

Test statistic:

$$F_0 = \frac{S_1^2}{S_2^2} \tag{5.13}$$

Under H_0, $F_0 \sim F_{n_1-1,\, n_2-1}$.

Convention (Procedure 5.1): Label the populations so the **larger** sample variance sits in the numerator. Then $F_0 \geq 1$ and the two-tailed test needs only the upper critical value.

F-Test for σ_1^2/σ_2^2

Procedure 5.1

Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2 \qquad (\text{equivalently } H_0 : \sigma_1^2/\sigma_2^2 = 1)$$

Test statistic:

$$F_0 = \frac{S_1^2}{S_2^2} \tag{5.13}$$

Under H_0, $F_0 \sim F_{n_1-1,\, n_2-1}$.

Convention (Procedure 5.1): Label the populations so the **larger** sample variance sits in the numerator. Then $F_0 \geq 1$ and the two-tailed test needs only the upper critical value.

Assumption: both populations **normal**. The F test is sensitive to this

Decision Rules for the F-Test

Critical values from Table B.5 and Table B.6

Let $u = n_1 - 1$, $v = n_2 - 1$.

Alternative H_1	Reject H_0 if
$\sigma_1^2 \neq \sigma_2^2$	$F_0 > f_{\alpha/2,\,u,\,v}$ or $F_0 < f_{1-\alpha/2,\,u,\,v}$

Decision Rules for the F-Test

Critical values from Table B.5 and Table B.6

Let $u = n_1 - 1$, $v = n_2 - 1$.

Alternative H_1	**Reject** H_0 **if**
$\sigma_1^2 \neq \sigma_2^2$	$F_0 > f_{\alpha/2,\, u,\, v}$ or $F_0 < f_{1-\alpha/2,\, u,\, v}$
$\sigma_1^2 > \sigma_2^2$	$F_0 > f_{\alpha,\, u,\, v}$

Decision Rules for the F-Test

Critical values from Table B.5 and Table B.6

Let $u = n_1 - 1$, $v = n_2 - 1$.

Alternative H_1	**Reject** H_0 **if**
$\sigma_1^2 \neq \sigma_2^2$	$F_0 > f_{\alpha/2,\,u,\,v}$ or $F_0 < f_{1-\alpha/2,\,u,\,v}$
$\sigma_1^2 > \sigma_2^2$	$F_0 > f_{\alpha,\,u,\,v}$
$\sigma_1^2 < \sigma_2^2$	$F_0 < f_{1-\alpha,\,u,\,v}$

Decision Rules for the F-Test

Critical values from Table B.5 and Table B.6

Let $u = n_1 - 1$, $v = n_2 - 1$.

Alternative H_1	**Reject H_0 if**
$\sigma_1^2 \neq \sigma_2^2$	$F_0 > f_{\alpha/2,\,u,\,v}$ or $F_0 < f_{1-\alpha/2,\,u,\,v}$
$\sigma_1^2 > \sigma_2^2$	$F_0 > f_{\alpha,\,u,\,v}$
$\sigma_1^2 < \sigma_2^2$	$F_0 < f_{1-\alpha,\,u,\,v}$

Lower-tail values from an upper-tail table — the book's reciprocal identity:

$$f_{1} = \frac{1}{\quad} \tag{5.12}$$

Confidence Interval for σ_1^2/σ_2^2

A $100(1-\alpha)\%$ CI for σ_1^2/σ_2^2:

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{f_{\alpha/2,\, n_1-1,\, n_2-1}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} \cdot f_{\alpha/2,\, n_2-1,\, n_1-1} \tag{5.14}$$

Confidence Interval for σ_1^2/σ_2^2

A $100(1-\alpha)\%$ CI for σ_1^2/σ_2^2:

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{f_{\alpha/2,\, n_1-1,\, n_2-1}} \;\le\; \frac{\sigma_1^2}{\sigma_2^2} \;\le\; \frac{s_1^2}{s_2^2} \cdot f_{\alpha/2,\, n_2-1,\, n_1-1} \tag{5.14}$$

Note: The two F critical values have **swapped** degrees of freedom (a consequence of (5.12)).

Confidence Interval for σ_1^2/σ_2^2

A $100(1-\alpha)\%$ CI for σ_1^2/σ_2^2:

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{f_{\alpha/2,\, n_1-1,\, n_2-1}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} \cdot f_{\alpha/2,\, n_2-1,\, n_1-1} \tag{5.14}$$

Note: The two F critical values have **swapped** degrees of freedom (a consequence of (5.12)).

If this interval contains 1, the data is consistent with $\sigma_1^2 = \sigma_2^2$.

Example: Consistency of Two Checkout Flows

Example 5.6 in the textbook uses the same recipe

A startup A/B tests two checkout flows and worries about **consistency**, not just speed. Completion times (minutes):

- Flow A: $n_1 = 16$, $s_1^2 = 0.85$ (min^2)
- Flow B: $n_2 = 21$, $s_2^2 = 0.32$ (min^2)

Test whether the variances differ at $\alpha = 0.05$.

Example: Consistency of Two Checkout Flows

Example 5.6 in the textbook uses the same recipe

A startup A/B tests two checkout flows and worries about **consistency**, not just speed. Completion times (minutes):

- Flow A: $n_1 = 16$, $s_1^2 = 0.85$ (min^2)
- Flow B: $n_2 = 21$, $s_2^2 = 0.32$ (min^2)

Test whether the variances differ at $\alpha = 0.05$.

Steps 1–3: Parameter: σ_1^2/σ_2^2. F-test (5.13), larger variance already labeled 1. $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$, $\alpha = 0.05$.

Example: Consistency of Two Checkout Flows

Example 5.6 in the textbook uses the same recipe

A startup A/B tests two checkout flows and worries about **consistency**, not just speed. Completion times (minutes):

- Flow A: $n_1 = 16$, $s_1^2 = 0.85$ (min^2)
- Flow B: $n_2 = 21$, $s_2^2 = 0.32$ (min^2)

Test whether the variances differ at $\alpha = 0.05$.

Steps 1–3: Parameter: σ_1^2/σ_2^2. F-test (5.13), larger variance already labeled 1. $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$, $\alpha = 0.05$.

Step 6: $F_0 = \dfrac{s_1^2}{s_2^2} = \dfrac{0.85}{0.32} = 2.656, \quad u = 15,\ v = 20.$

Example: Consistency of Two Checkout Flows

Example 5.6 in the textbook uses the same recipe

A startup A/B tests two checkout flows and worries about **consistency**, not just speed. Completion times (minutes):

- Flow A: $n_1 = 16$, $s_1^2 = 0.85$ (min^2)
- Flow B: $n_2 = 21$, $s_2^2 = 0.32$ (min^2)

Test whether the variances differ at $\alpha = 0.05$.

Steps 1–3: Parameter: σ_1^2/σ_2^2. F-test (5.13), larger variance already labeled 1. $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$, $\alpha = 0.05$.

Step 6: $F_0 = \dfrac{s_1^2}{s_2^2} = \dfrac{0.85}{0.32} = 2.656, \quad u = 15,\ v = 20.$

Steps 4–5, 7: $f_{0.025,\,15,\,20} = 2.57$. Since $F_0 = 2.656 > 2.57$: **reject H_0**. Flow A's completion times are significantly more variable than Flow B's.

Example: Checkout Flows (Confidence Interval)

95% CI for σ_1^2/σ_2^2 via (5.14), with $f_{0.025,\,15,\,20} = 2.57$ and $f_{0.025,\,20,\,15} = 2.76$:

Example: Checkout Flows (Confidence Interval)

95% CI for σ_1^2/σ_2^2 via (5.14), with $f_{0.025,\,15,\,20} = 2.57$ and $f_{0.025,\,20,\,15} = 2.76$:

$$\text{Lower} = \frac{0.85}{0.32} \cdot \frac{1}{2.57} = \frac{2.656}{2.57} = 1.033$$

Example: Checkout Flows (Confidence Interval)

95% CI for σ_1^2/σ_2^2 via (5.14), with $f_{0.025,\,15,\,20} = 2.57$ and $f_{0.025,\,20,\,15} = 2.76$:

$$\text{Lower} = \frac{0.85}{0.32} \cdot \frac{1}{2.57} = \frac{2.656}{2.57} = 1.033$$

$$\text{Upper} = \frac{0.85}{0.32} \cdot 2.76 = 2.656 \times 2.76 = 7.331$$

Example: Checkout Flows (Confidence Interval)

95% CI for σ_1^2/σ_2^2 via (5.14), with $f_{0.025,\,15,\,20} = 2.57$ and $f_{0.025,\,20,\,15} = 2.76$:

$$\text{Lower} = \frac{0.85}{0.32} \cdot \frac{1}{2.57} = \frac{2.656}{2.57} = 1.033$$

$$\text{Upper} = \frac{0.85}{0.32} \cdot 2.76 = 2.656 \times 2.76 = 7.331$$

95% CI: $(1.033,\ 7.331)$. The interval **excludes** 1, consistent with rejecting H_0.

Example: Checkout Flows (Confidence Interval)

95% CI for σ_1^2/σ_2^2 via (5.14), with $f_{0.025,\,15,\,20} = 2.57$ and $f_{0.025,\,20,\,15} = 2.76$:

$$\text{Lower} = \frac{0.85}{0.32} \cdot \frac{1}{2.57} = \frac{2.656}{2.57} = 1.033$$

$$\text{Upper} = \frac{0.85}{0.32} \cdot 2.76 = 2.656 \times 2.76 = 7.331$$

95% CI: $(1.033,\ 7.331)$. The interval **excludes** 1, consistent with rejecting H_0.

Practical reading: Flow A's variance is plausibly anywhere from about equal to 7 times Flow B's — B is the more predictable experience.

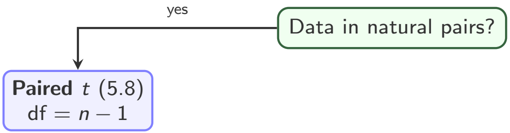
Choosing Among the Two-Sample Tools

Procedure 5.2; see Figure 5.6

Data in natural pairs?

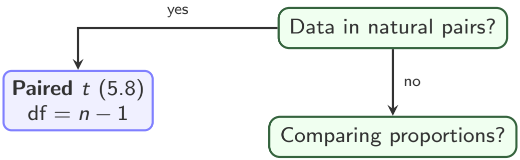
Choosing Among the Two-Sample Tools

Procedure 5.2; see Figure 5.6



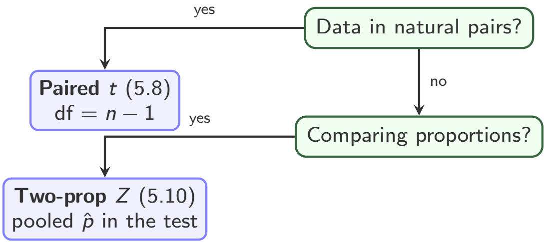
Choosing Among the Two-Sample Tools

Procedure 5.2; see Figure 5.6



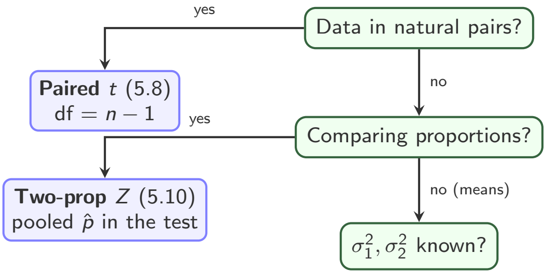
Choosing Among the Two-Sample Tools

Procedure 5.2; see Figure 5.6



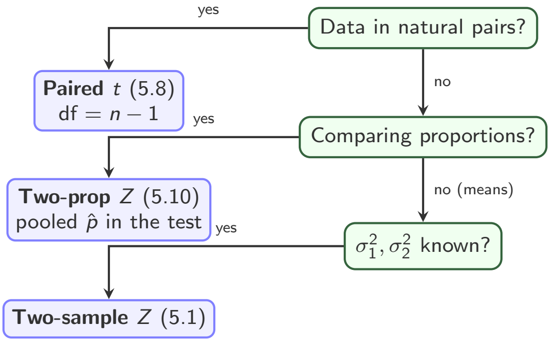
Choosing Among the Two-Sample Tools

Procedure 5.2; see Figure 5.6



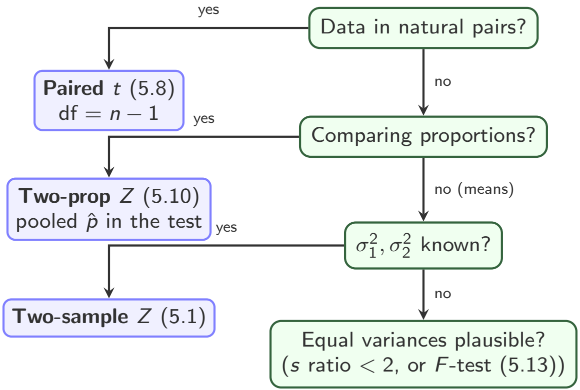
Choosing Among the Two-Sample Tools

Procedure 5.2; see Figure 5.6



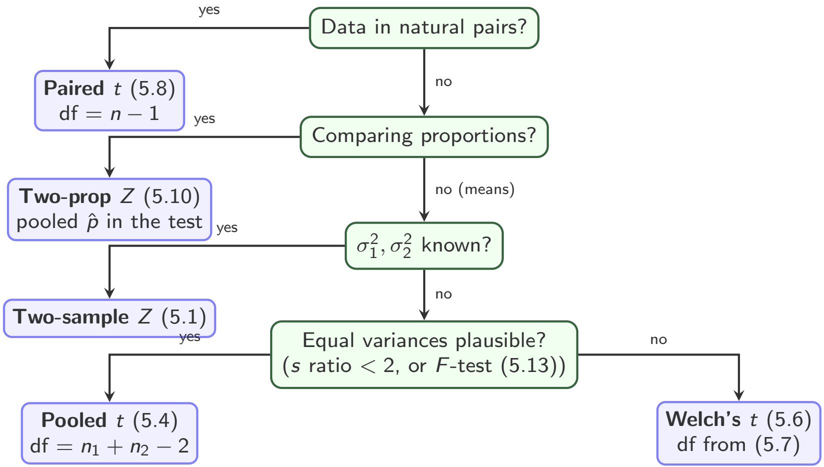
Choosing Among the Two-Sample Tools

Procedure 5.2; see Figure 5.6



Choosing Among the Two-Sample Tools

Procedure 5.2; see Figure 5.6



Complete Summary: Which Test to Use?

The book's master table is Table 5.6

Parameter	Conditions	Test Statistic	Distribution	Book Eq.
Single sample (Ch. 3–4):				
μ	σ known	$Z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$N(0,1)$	(4.3)
μ	σ unknown	$T_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	t_{n-1}	(4.6)
σ^2	Normal pop.	$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$	χ_{n-1}^2	(4.7)
p	Large n	Z_0	$N(0,1)$	(4.8)
Two samples (Ch. 5):				
$\mu_1 - \mu_2$	σ_1, σ_2 known	Z_0	$N(0,1)$	(5.1)
$\mu_1 - \mu_2$	$\sigma_1^2 = \sigma_2^2$ unkn.	T_0 (pooled)	$t_{n_1+n_2-2}$	(5.4)
$\mu_1 - \mu_2$	$\sigma_1^2 \neq \sigma_2^2$ unkn.	T_0 (Welch)	t_ν, (5.7)	(5.6)
μ_D	Paired data	$T_0 = \frac{\bar{d}}{s_D/\sqrt{n}}$	t_{n-1}	(5.8)
σ_1^2/σ_2^2	Normal pops.	$F_0 = s_1^2/s_2^2$	F_{n_1-1,n_2-1}	(5.13)

Common Mistakes to Avoid (Updated)

- **Using an independent test for paired data** (or vice versa).
  Ask: are the observations naturally linked? If yes $\Rightarrow$ paired (5.8).

Common Mistakes to Avoid (Updated)

- **Using an independent test for paired data** (or vice versa).
  Ask: are the observations naturally linked? If yes $\Rightarrow$ paired (5.8).
- **Wrong df for Welch's test.**
  Use the Welch–Satterthwaite formula (5.7) and round **down** — *not* $n_1 + n_2 - 2$.

Common Mistakes to Avoid (Updated)

- **Using an independent test for paired data** (or vice versa).
  Ask: are the observations naturally linked? If yes $\Rightarrow$ paired (5.8).
- **Wrong df for Welch's test.**
  Use the Welch–Satterthwaite formula (5.7) and round **down** — *not* $n_1 + n_2 - 2$.
- **F-test: swapping numerator and denominator df.**
  $F_{u,v}$: u = numerator df ($n_1 - 1$), v = denominator df ($n_2 - 1$).

Common Mistakes to Avoid (Updated)

- **Using an independent test for paired data** (or vice versa).
  Ask: are the observations naturally linked? If yes $\Rightarrow$ paired (5.8).
- **Wrong df for Welch's test.**
  Use the Welch–Satterthwaite formula (5.7) and round **down** — *not* $n_1 + n_2 - 2$.
- **F-test: swapping numerator and denominator df.**
  $F_{u,v}$: u = numerator df $(n_1 - 1)$, v = denominator df $(n_2 - 1)$.
- **Two-proportion CI with the pooled $\hat{p}$.**
  Pooled $\hat{p}$ is for the test (5.10) only; the CI (5.11) uses $\hat{p}_1, \hat{p}_2$.

Common Mistakes to Avoid (Updated)

- **Using an independent test for paired data** (or vice versa).
  Ask: are the observations naturally linked? If yes $\Rightarrow$ paired (5.8).
- **Wrong df for Welch's test.**
  Use the Welch–Satterthwaite formula (5.7) and round **down** — *not* $n_1 + n_2 - 2$.
- **F-test: swapping numerator and denominator df.**
  $F_{u,v}$: $u =$ numerator df $(n_1 - 1)$, $v =$ denominator df $(n_2 - 1)$.
- **Two-proportion CI with the pooled $\hat{p}$.**
  Pooled $\hat{p}$ is for the test (5.10) only; the CI (5.11) uses $\hat{p}_1, \hat{p}_2$.
- **Skipping the equal-variance check.**
  Compare s_1/s_2 (ratio $\geq 2 \Rightarrow$ Welch) or run the F-test (5.13) first.

Lecture 12 Summary

- **Welch's t-test** (5.6): unequal variances; keep s_1^2, s_2^2 separate; df from (5.7), rounded down.

Lecture 12 Summary

- **Welch's t-test** (5.6): unequal variances; keep s_1^2, s_2^2 separate; df from (5.7), rounded down.
- **Paired t-test** (5.8): compute differences d_j, then treat as one sample; CI (5.9); df $= n - 1$.

Lecture 12 Summary

- **Welch's t-test** (5.6): unequal variances; keep s_1^2, s_2^2 separate; df from (5.7), rounded down.
- **Paired t-test** (5.8): compute differences d_j, then treat as one sample; CI (5.9); df $= n - 1$.
- **Two-proportion Z-test** (5.10): pool $\hat{p}$ under H_0; the CI (5.11) uses individual $\hat{p}_1, \hat{p}_2$.

Lecture 12 Summary

- **Welch's t-test** (5.6): unequal variances; keep s_1^2, s_2^2 separate; df from (5.7), rounded down.
- **Paired t-test** (5.8): compute differences d_j, then treat as one sample; CI (5.9); df $= n - 1$.
- **Two-proportion Z-test** (5.10): pool $\hat{p}$ under H_0; the CI (5.11) uses individual $\hat{p}_1, \hat{p}_2$.
- **F-test** (5.13): compares σ_1^2/σ_2^2; reciprocal identity (5.12); CI (5.14).

Lecture 12 Summary

- **Welch's t-test** (5.6): unequal variances; keep s_1^2, s_2^2 separate; df from (5.7), rounded down.
- **Paired t-test** (5.8): compute differences d_j, then treat as one sample; CI (5.9); df $= n - 1$.
- **Two-proportion Z-test** (5.10): pool $\hat{p}$ under H_0; the CI (5.11) uses individual $\hat{p}_1, \hat{p}_2$.
- **F-test** (5.13): compares σ_1^2/σ_2^2; reciprocal identity (5.12); CI (5.14).
- **Choosing:** follow Procedure 5.2 — paired? proportions? σ's known? equal variances?

Lecture 12 Summary

- **Welch's t-test** (5.6): unequal variances; keep s_1^2, s_2^2 separate; df from (5.7), rounded down.
- **Paired t-test** (5.8): compute differences d_j, then treat as one sample; CI (5.9); df $= n - 1$.
- **Two-proportion Z-test** (5.10): pool $\hat{p}$ under H_0; the CI (5.11) uses individual $\hat{p}_1, \hat{p}_2$.
- **F-test** (5.13): compares σ_1^2/σ_2^2; reciprocal identity (5.12); CI (5.14).
- **Choosing:** follow Procedure 5.2 — paired? proportions? σ's known? equal variances?
- The seven-step procedure and the CI–test duality carry through everything.

Class Announcements

- **Major Exam 1 covers Chapters 2–5** (through today's material)

Class Announcements

- **Major Exam 1 covers Chapters 2–5** (through today's material)
- Chapter 2 supplies the background sampling distributions for single- and two-sample statistics

Class Announcements

- **Major Exam 1 covers Chapters 2–5** (through today's material)
- Chapter 2 supplies the background sampling distributions for single- and two-sample statistics
- Next two lectures: full review (Lecture 13) and worked practice problems (Lecture 14)