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 (Ch. 10)

Lecture 12 Outline

- Quick review of Lecture 11 (two-sample Z and pooled t)
- Welch's t-test: unequal variances (Sec. 10-2.2)
- The paired t-test (Sec. 10-4)
- F-test for the ratio of two variances (Sec. 10-5)
- Inference on two population proportions (Sec. 10-6)

Review: What We Covered in Lecture 11

Two-sample Z-test (σ_1, σ_2 known):

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

Pooled t-test ($\sigma_1^2 = \sigma_2^2$ assumed, both unknown):

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

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

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.

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

The distribution for T_0 is approximately t_ν, where ν of degrees of freedom is computed from the **Welch–Satterthwaite** formula.

Welch–Satterthwaite Degrees of Freedom

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

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

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

Comparison of standard errors vs. pooled t test:

Method	**Standard Error**
Pooled t	$s_p \sqrt{1/n_1 + 1/n_2}$
Welch's t	$\sqrt{s_1^2/n_1 + s_2^2/n_2}$

Example: Comparing Catalyst Yields

A chemical engineer tests two catalysts for reaction yield (%).

- Catalyst A: $n_1 = 12$, $\bar{x}_1 = 85.6$, $s_1 = 3.2$
- Catalyst B: $n_2 = 15$, $\bar{x}_2 = 81.9$, $s_2 = 6.8$

Do the mean yields differ at $\alpha = 0.05$?

Additional step: Check s ratio: $s_2/s_1 = 6.8/3.2 = 2.125 > 2$. Do **not** assume equal variances.

Step 1: Parameter: $\mu_1 - \mu_2$. Variances unknown and unequal. Use Welch's t-test.

Step 2: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 \neq 0$, $\alpha = 0.05$.

Example: Catalyst Yields (Solution)

Step 3: Standard error: $\sqrt{\dfrac{3.2^2}{12} + \dfrac{6.8^2}{15}} = \sqrt{0.8533 + 3.0827} = \sqrt{3.9360} = 1.984$

$$T_0 = \frac{(85.6 - 81.9) - 0}{1.984} = \frac{3.7}{1.984} = 1.865$$

Degrees of freedom (Welch–Satterthwaite):

$$\nu = \frac{(0.8533 + 3.0827)^2}{\dfrac{0.8533^2}{11} + \dfrac{3.0827^2}{14}} = \frac{15.492}{0.0662 + 0.6789} = \frac{15.492}{0.7451} = 20.79 \;\Rightarrow\; \nu = 20$$

Steps 4–6: $t_{0.025,\,20} = 2.086$. $|T_0| = 1.865 < 2.086$.

Fail to reject H_0. Not enough evidence to conclude the catalysts differ.

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
- Two methods applied to the **same** specimen
- Matched subjects (twin studies, left/right comparisons)

Key idea: Do not use two samples, analyze them as a *single sample* by computing the **differences** $D_j = X_{1j} - X_{2j}$

This removes subject-to-subject variability, making the test **more powerful**.

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:

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

Under H_0, $T_0 \sim t_{n-1}$.

CI for μ_D: $\bar{d} \pm t_{\alpha/2,\,n-1} \dfrac{s_D}{\sqrt{n}}$

Example: Fuel Efficiency Before and After Tune-Up

Eight fleet vehicles are tested for fuel efficiency (km/L) before and after an engine tune-up.

Vehicle	1	2	3	4	5	6	7	8
Before (x_{1j})	12.1	14.3	11.8	13.5	15.2	12.7	14.0	13.1
After (x_{2j})	12.9	14.8	12.5	14.1	15.6	13.4	14.9	13.6
$d_j = x_{2j} - x_{1j}$	0.8	0.5	0.7	0.6	0.4	0.7	0.9	0.5

Does the tune-up improve fuel efficiency? Use $\alpha = 0.05$.

Note: Paired data (same vehicle before and after). Use paired t-test.
Let $d_j =$ After − Before. $H_0 : \mu_D = 0$ vs $H_1 : \mu_D > 0$ (right-tailed).

Example: Fuel Efficiency (Solution)

Step 3: From the differences: $\bar{d} = \dfrac{0.8 + 0.5 + 0.7 + 0.6 + 0.4 + 0.7 + 0.9 + 0.5}{8} = \dfrac{5.1}{8} = 0.6375$

$$s_D = \sqrt{\frac{\sum(d_j - \bar{d})^2}{n-1}} = \sqrt{\frac{0.1888}{7}} = \sqrt{0.02697} = 0.1642$$

$$T_0 = \frac{0.6375 - 0}{0.1642/\sqrt{8}} = \frac{0.6375}{0.05806} = 10.98$$

Steps 4–6: $\nu = 7$. $t_{0.05,7} = 1.895$. $T_0 = 10.98 \gg 1.895$. P-value ≈ 0.
Reject H_0. Strong evidence that the tune-up improves fuel efficiency.

95% lower bound on μ_D: $0.6375 - 1.895(0.05806) = 0.528$ km/L.

Paired vs. Independent: How to Decide

	Two-Sample	**Paired**
Design	Two separate groups	Same subjects, 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	$n - 1$
Advantage	Simpler design	No subject variability

Rule of thumb: If the problem says "same," "matched," "before/after," or gives

Comparing Two Variances: The F-Test

Section 10-5

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

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

The F distribution: If $S_1^2 \sim \chi^2_{n_1-1}/(n_1-1)$ and $S_2^2 \sim \chi^2_{n_2-1}/(n_2-1)$ independently, then

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

The F distribution has **two** degree-of-freedom parameters: the numerator df ($u = n_1 - 1$) and the denominator df ($v = n_2 - 1$).

F-Test for σ_1^2/σ_2^2

Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

(equivalently: $H_0 : \sigma_1^2/\sigma_2^2 = 1$)

Test statistic:

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

Under H_0 ($\sigma_1^2 = \sigma_2^2$), $F_0 \sim F_{n_1-1,\, n_2-1}$.

Convention: Place the **larger** sample variance in the numerator.
This ensures $F_0 \geq 1$ and simplifies the two-tailed test.

Decision Rules for F-Test

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 $F_0 < f_{1-\alpha/2,\, u,\, v}$
$\sigma_1^2 > \sigma_2^2$	$F_0 > f_{\alpha,\, u,\, v}$
$\sigma_1^2 < \sigma_2^2$	$F_0 < f_{1-\alpha,\, u,\, v}$

Useful identity: $f_{1-\alpha,\, u,\, v} = \dfrac{1}{f_{\alpha,\, v,\, u}}$ (let's you compute the lower critical value from the upper-tail table).

Confidence Interval for σ_1^2/σ_2^2

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

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

Note: The two F critical values have **swapped** degrees of freedom.

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

Example: Comparing Variability of Two Suppliers

An ISE engineer receives steel rods from two suppliers and measures the diameter:

- Supplier 1: $n_1 = 16$, $s_1^2 = 0.042$ (mm^2)
- Supplier 2: $n_2 = 21$, $s_2^2 = 0.018$ (mm^2)

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

Step 1: Parameter: σ_1^2/σ_2^2. Use F-test.
Step 2: $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$, $\alpha = 0.05$.

Step 3: $F_0 = \dfrac{s_1^2}{s_2^2} = \dfrac{0.042}{0.018} = 2.333$

$u = n_1 - 1 = 15$, $v = n_2 - 1 = 20$.

Steps 4–6: $f_{0.025,15,20} = 2.57$ (from F-table). $F_0 = 2.333 < 2.57$. **Fail to reject H_0.** Not enough evidence that the variances differ.

Inference on Two Population Proportions

Section 10-6

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:

$$H_0 : p_1 = p_2 \quad \text{vs} \quad H_1 : p_1 \neq p_2$$

(equivalently: $H_0 : p_1 - p_2 = 0$)

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

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

Test Statistic for Two Proportions

Test statistic:

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

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).

Decision rules are identical to the standard Z-test:
Two-tailed: reject if $|Z_0| > z_{\alpha/2}$.
Right-tailed: reject if $Z_0 > z_\alpha$.

Confidence Interval for $p_1 - p_2$

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

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

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

Condition: Both samples should be "large enough" so that $n_i\hat{p}_i \geq 5$ and $n_i(1-\hat{p}_i) \geq 5$.

Example: Weld Defect Rates

In a pipeline construction project, an engineer compares the defect rates of welds from two crews.

- Crew A: $n_1 = 200$ welds inspected, $x_1 = 18$ defective, $\hat{p}_1 = 0.09$
- Crew B: $n_2 = 250$ welds inspected, $x_2 = 32$ defective, $\hat{p}_2 = 0.128$

Is there a difference in defect rates? Use $\alpha = 0.05$.

Step 1–2: $H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$, $\alpha = 0.05$. Use two-proportion Z-test.

Step 3: Pooled proportion: $\hat{p} = \dfrac{18 + 32}{200 + 250} = \dfrac{50}{450} = 0.1111$

$$Z_0 = \frac{0.09 - 0.128}{\sqrt{0.1111(0.8889)\left(\frac{1}{200} + \frac{1}{250}\right)}} = \frac{-0.038}{\sqrt{0.0009378}} = \frac{-0.038}{0.03062} = -1.241$$

Example: Weld Defect Rates (Conclusion)

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

Step 5: $|Z_0| = 1.241 < 1.96$. Not in the rejection region.
P-value $= 2[1 - \Phi(1.241)] = 2(1 - 0.8926) = 2(0.1074) = 0.2148$

Step 6: Fail to reject H_0. At $\alpha = 0.05$, there is not sufficient evidence that the weld defect rates differ between the two crews.

95% CI for $p_1 - p_2$:

$$
\begin{aligned}
&= (0.09 - 0.128) \pm 1.96\sqrt{\frac{0.09(0.91)}{200} + \frac{0.128(0.872)}{250}} \\
&= -0.038 \pm 1.96(0.02993) = -0.038 \pm 0.0587 \\
&= (-0.097,\ 0.021) \text{ (contains 0, consistent with failing to reject).}
\end{aligned}
$$

Complete Summary: Which Test to Use?

Parameter	Conditions	Test Statistic	Distribution
Single sample (Ch. 8–9):			
μ	σ known	$Z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$N(0,1)$
μ	σ unknown	$T_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	t_{n-1}
σ^2	Normal pop.	$\chi^2_0 = \frac{(n-1)s^2}{\sigma^2_0}$	χ^2_{n-1}
p	Large n	Z_0	$N(0,1)$
Two samples (Ch. 10):			
$\mu_1 - \mu_2$	σ_1, σ_2 known	Z_0	$N(0,1)$
$\mu_1 - \mu_2$	$\sigma^2_1 = \sigma^2_2$ unkn.	T_0 (pooled)	$t_{n_1+n_2-2}$
$\mu_1 - \mu_2$	$\sigma^2_1 \neq \sigma^2_2$ unkn.	T_0 (Welch)	t_ν (W-S)
μ_D	Paired data	$T_0 = \frac{\bar{d}}{s_D/\sqrt{n}}$	t_{n-1}
σ^2_1/σ^2_2	Normal pops.	$F_0 = s^2_1/s^2_2$	F_{n_1-1,n_2-1}
$p_1 - p_2$	Large n_1, n_2	Z_0	$N(0,1)$

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.
- **Wrong df for Welch's test.**
  Must use the Welch–Satterthwaite formula, *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: using pooled $\hat{p}$ instead of $\hat{p}_1, \hat{p}_2$.**
  Pooled $\hat{p}$ is for the test only; CI uses individual proportions.
- **Forgetting to check the equal-variance assumption.**
  Use the F-test or compare s_1/s_2. If ratio > 2, use Welch.

Lecture 12 Summary

- **Welch's t-test:** For unequal variances. Keep s_1^2, s_2^2 separate. Degrees of freedom from the Welch–Satterthwaite formula (round down).
- **Paired t-test:** Compute differences d_j, then treat as a one-sample problem. $T_0 = \bar{d}/(s_D/\sqrt{n})$, with $n-1$ df.
- **F-test:** Compares σ_1^2/σ_2^2. $F_0 = s_1^2/s_2^2 \sim F_{n_1-1,\, n_2-1}$. Not symmetric: numerator and denominator df matter.
- **Two-proportion Z-test:** Pool the proportions under H_0. CI uses individual $\hat{p}_1, \hat{p}_2$ (not pooled).
- The **6-step procedure** and **CI-test connection** remain the same throughout.

Class Announcements

- Major 1 Exam will cover Chapters 8-10 (what we cover today)
- Chapter 7 mostly provides background formula (for single sample and two sample statistics)
- Best of luck for Major 1! We will have a final practice quiz next week.