Lectures 8,9
Testing Hypotheses for the Difference Between Two Population Means

Prof. Dr Saad Abed Madhi

1. Introduction

In many medical, biological, or social science applications, researchers are interested in comparing **two population means**. Examples include comparing:

- Mean blood pressure between treatment and control groups
- Mean hemoglobin between males and females
- Mean recovery times under two protocols
- Mean birthweights between two regions

The statistical tool used is **hypothesis testing for $\mu_1 - \mu_2$**, the difference between two means.

We cover the following cases:

1. **Independent samples, known variances (Z-test)**
2. **Independent samples, unknown but equal variances (pooled t-test)**
3. **Independent samples, unknown and unequal variances (Welch's t-test)**
4. **Paired samples (paired t-test)**

2. General Structure of Hypothesis Testing

Every hypothesis test follows these steps:

Step 1: State the hypotheses

Let μ_1 and μ_2 be the two population means.

Most common hypotheses:

- **Two-sided test**

  $H_0 : \mu_1 - \mu_2 = 0$

  $H_a : \mu_1 - \mu_2 \neq 0$
- **Right-tailed**

  $H_0 : \mu_1 - \mu_2 = 0$

  $H_a : \mu_1 - \mu_2 > 0$
- **Left-tailed**

  $H_0 : \mu_1 - \mu_2 = 0$

  $H_a : \mu_1 - \mu_2 < 0$

Step 2: Compute test statistic

This depends on whether variances are known, equal, unequal, or paired.

Step 3: Determine critical value(s)

- Use standard normal (Z) distribution when σ's are known.
- Use t-distribution when σ's are unknown.

Step 4: Decision rule

Compare:

$$\text{Test statistic} \quad \text{vs.} \quad \text{Critical value}$$

Reject or fail to reject H_0.

**3. Case 1: Independent Samples, Variances Known

(Z-Test for Two Means)**

This case is rare in medical studies but is the simplest.

Test Statistic

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Usually, $(\mu_1 - \mu_2)_0 = 0$.

Critical Regions

At significance α:

- Two-sided: Reject if $|Z| > Z_{\alpha/2}$
- Right-tailed: Reject if $Z > Z_{\alpha}$
- Left-tailed: Reject if $Z < -Z_{\alpha}$

Typical Z-values:

α	Two-sided	One-tailed
0.10	±1.645	1.28
0.05	±1.96	1.645
0.01	±2.576	2.33

**4. Case 2: Independent Samples, Variances Unknown but Equal

(Pooled Two-Sample t-Test)**

This is widely used in clinical trials where two groups are assumed to have similar variability.

Assumptions

- Samples independent
- Variances equal: $\sigma_1^2 = \sigma_2^2$

Pooled Variance

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Test Statistic

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Degrees of Freedom

$$df = n_1 + n_2 - 2$$

Critical Values

Use t distribution with df above.

Decision rules:

- Two-sided: Reject if $|T| > t_{\alpha/2,df}$
- Right-tailed: Reject if $T > t_{\alpha,df}$
- Left-tailed: Reject if $T < -t_{\alpha,df}$

**5. Case 3: Independent Samples, Variances Unknown and Unequal

(Welch's t-test)**

Most realistic for biostatistics.

Test Statistic

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

Degrees of Freedom (Welch–Satterthwaite)

$$df = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Use the resulting df in t tables.

Decision rules same as above.

6. Case 4: Paired Samples (Paired t-Test)

Used for:

- Before/after studies
- Right/left eye measurements
- Same patient measured twice
- Cross-over trials

Instead of comparing two means independently, analyze **differences.**

Let differences be $d_i = X_{1i} - X_{2i}$

We test:

$$H_0 : \mu_d = 0$$

Test Statistic

$$T = \frac{\bar{d}}{s_d/\sqrt{n}}$$

Where:

- $\bar{d}$ = mean of the differences
- s_d = standard deviation of the differences

Degrees of Freedom

$$df = n - 1$$

Decision Rules

Use t critical values:

- Two-sided: Reject if $|T| > t_{\alpha/2,df}$
- One-sided: Reject in the direction of Ha

7. Choosing the Correct Test

Scenario	Variances Known	Variances Equal (Unknown)	Variances Unequal (Unknown)	Paired Data
Test	Two-sample Z	Pooled t	Welch t	Paired t

8. Summary of Critical Value Approach

You ONLY compare:

$$\boxed{\text{Test Statistic} \quad \text{vs.} \quad \text{Critical Value}}$$

Decision rule:

- If statistic falls in rejection region → **Reject H_0**
- Otherwise → **Fail to reject H_0**

9. Worked Example (One Full Example)

(Just one here—if you want 10, I will produce them.)

Example: Two Independent Samples, Equal Variances

A study compares systolic BP between two groups.

Group	n	$\bar{x}$	s
Treatment	30	128	10
Control	32	135	12

Test at $\alpha=0.05$ if treatment reduces mean BP

(two-sided test).

Step 1: Hypotheses

$$H_0 : \mu_1 - \mu_2 = 0, \quad H_a : \mu_1 - \mu_2 \neq 0$$

Step 2: Compute pooled variance

$$s_p^2 = \frac{(29)(10^2) + (31)(12^2)}{60} = \frac{2900 + 4464}{60}$$

$$s_p^2 = \frac{7364}{60} = 122.73$$

$$s_p = 11.07$$

Step 3: Test statistic

$$T = \frac{128 - 135}{11.07\sqrt{\frac{1}{30} + \frac{1}{32}}}$$

Compute denominator:

$$\frac{1}{30} + \frac{1}{32} = 0.0646$$

$$SE = 11.07\sqrt{0.0646} = 11.07(0.254) = 2.81$$

$$T = \frac{-7}{2.81} = -2.49$$

Step 4: Critical value (two-sided, df=60, α=0.05)

$$t_{0.025,60} \approx 2.000$$

Step 5: Decision

$$|T| = 2.49 > 2.000 = t_{\text{critical}}$$

→ Reject H_0

Interpretation

There is significant evidence (critical value method) that the two population means are different; treatment reduces BP.

Detailed worked examples

1. Z-Test, Independent Samples, Known Variances (Two-Sided)

Data:

- Group 1: $n_1 = 40$, $\bar{x}_1 = 105$, $\sigma_1 = 10$
- Group 2: $n_2 = 50$, $\bar{x}_2 = 100$, $\sigma_2 = 12$

Hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_a : \mu_1 - \mu_2 \neq 0$$

$\alpha = 0.05$

Step 1: Compute SE

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{100}{40} + \frac{144}{50}} = \sqrt{2.5 + 2.88} = \sqrt{5.38} \approx 2.32$$

Step 2: Test Statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{SE} = \frac{105 - 100}{2.32} = \frac{5}{2.32} \approx 2.155$$

Step 3: Critical Value (Two-Sided, α=0.05)

$$Z_{0.975} = 1.96$$

Step 4: Decision

$|Z| = 2.155 > 1.96 \Rightarrow \text{Reject } H_0$

Conclusion: Evidence suggests $\mu_1 \neq \mu_2$.

2. Z-Test, Independent Samples, Right-Tailed

Data:

- Group 1: $n_1 = 36$, $\bar{x}_1 = 52$, $\sigma_1 = 8$
- Group 2: $n_2 = 40$, $\bar{x}_2 = 48$, $\sigma_2 = 10$

Hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_a : \mu_1 - \mu_2 > 0$$

$\alpha = 0.05$

Step 1: SE

$$SE = \sqrt{\frac{8^2}{36} + \frac{10^2}{40}} = \sqrt{1.7778 + 2.5} = \sqrt{4.2778} \approx 2.07$$

Step 2: Z

$$Z = \frac{52 - 48}{2.07} = \frac{4}{2.07} \approx 1.933$$

Step 3: Critical Value

$$Z_{0.95} = 1.645$$

Step 4: Decision

$1.933 > 1.645$ → **Reject H_0**

Conclusion: Mean of Group 1 is significantly higher.

3. Pooled t-Test, Independent Samples, Unknown but Equal Variances (Two-Sided)

Data:

- Group 1: $n_1=25$, $\bar{x}_1=120$, $s_1=15$
- Group 2: $n_2=30$, $\bar{x}_2=110$, $s_2=12$

Hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_a : \mu_1 - \mu_2 \neq 0$$

$\alpha = 0.05$

Step 1: Pooled Variance

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{24*225+29*144}{53} = \frac{5400+4176}{53} = \frac{9576}{53} \approx 180.7$$

$$s_p = \sqrt{180.7} \approx 13.44$$

Step 2: SE

$$SE = s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 13.44\sqrt{\frac{1}{25}+\frac{1}{30}} = 13.44\sqrt{0.04+0.0333} = 13.44\sqrt{0.0733} \approx 13.44 * 0.271 \approx$$

Step 3: T statistic

$$T = \frac{120-110}{3.64} = \frac{10}{3.64} \approx 2.747$$

Step 4: Critical value (df=53, α=0.05, two-sided)

$$t_{0.975,53} \approx 2.006$$

Step 5: Decision

2.747 > 2.006 → **Reject H_0**

Conclusion: Significant difference in means.

4. Pooled t-Test, Independent Samples, Right-Tailed

Data:

- Group 1: $n_1=20$, $\bar{x}_1=78$, $s_1=10$
- Group 2: $n_2=22$, $\bar{x}_2=74$, $s_2=12$

Hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_a : \mu_1 - \mu_2 > 0$$

$\alpha = 0.05$

Step 1: Pooled Variance

$$s_p^2 = \frac{19 * 100 + 21 * 144}{40} = \frac{1900 + 3024}{40} = 4924/40 = 123.1$$

$$s_p = \sqrt{123.1} \approx 11.1$$

Step 2: SE

$$SE = 11.1\sqrt{\frac{1}{20}+\frac{1}{22}} = 11.1\sqrt{0.05+0.0455} = 11.1\sqrt{0.0955} \approx 11.1 * 0.309 \approx 3.43$$

Step 3: T statistic

$$T = \frac{78-74}{3.43} = 4/3.43 \approx 1.166$$

Step 4: Critical value (df=40, α=0.05, right-tail)

$$t_{0.95,40} \approx 1.684$$

Step 5: Decision

1.166 < 1.684 → **Do not reject H_0**

Conclusion: No significant evidence that $\mu_1 > \mu_2$.

5. Welch t-Test, Independent Samples, Unequal Variances (Two-Sided)

Data:

- Group 1: $n_1=15$, $\bar{x}_1=88$, $s_1=7$
- Group 2: $n_2=12$, $\bar{x}_2=95$, $s_2=10$

Hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_a : \mu_1 - \mu_2 \neq 0$$

$\alpha = 0.05$

Step 1: SE

$$SE = \sqrt{\frac{7^2}{15} + \frac{10^2}{12}} = \sqrt{\frac{49}{15} + \frac{100}{12}} = \sqrt{3.267 + 8.333} = \sqrt{11.6} \approx 3.406$$

Step 2: T statistic

$$T = \frac{88 - 95}{3.406} = \frac{-7}{3.406} \approx -2.056$$

Step 4: Critical value (two-sided, α=0.05, df=19)

$$t_{0.975,19} \approx 2.093$$

Step 5: Decision

$|T| = 2.056 < 2.093$ → **Do not reject H_0**

Conclusion: Insufficient evidence of a difference.

6. Welch t-Test, Independent Samples, Right-Tailed

Data:

- Group 1: $n_1=18$, $\bar{x}_1=112$, $s_1=15$
- Group 2: $n_2=16$, $\bar{x}_2=105$, $s_2=12$

Hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_a : \mu_1 - \mu_2 > 0$$

$\alpha=0.05$

Step 1: SE

$$SE = \sqrt{\frac{225}{18} + \frac{144}{16}} = \sqrt{12.5 + 9} = \sqrt{21.5} \approx 4.638$$

Step 2: T

$$T = \frac{112 - 105}{4.638} = 7/4.638 \approx 1.509$$

Step 3: df

$$df \approx \frac{21.5^2}{\frac{12.5^2}{17} + \frac{9^2}{15}} = \frac{462.25}{9.191 + 5.4} = \frac{462.25}{14.591} \approx 31.7 \approx 32$$

Step 4: Critical value (right-tail, df=32, α=0.05)

$$t_{0.95,32} \approx 1.694$$

Step 5: Decision

1.509 < 1.694 → **Do not reject H_0**

Conclusion: Not enough evidence that $\mu_1 > \mu_2$.

7. Paired t-Test, Two-Sided

Data: Before and after measurements on 12 patients:

Patient	Before	After	d = Before−After
1	120	115	5
2	130	128	2
3	125	122	3
...	...	...	...
n=12	$\bar{d} = 4$, s_d = 2.5		

Hypotheses:

$$H_0 : \mu_d = 0 \qquad H_a : \mu_d \neq 0$$

Step 1: SE

$$SE = \frac{s_d}{\sqrt{n}} = \frac{2.5}{\sqrt{12}} \approx 0.7217$$

Step 2: T

$$T = \frac{\bar{d}}{SE} = \frac{4}{0.7217} \approx 5.547$$

Step 3: Critical value (two-sided, df=11, α=0.05)

$$t_{0.975,11} \approx 2.201$$

Step 4: Decision

5.547 > 2.201 → **Reject H_0**

Conclusion: Treatment significantly reduces the measurement.

8. Paired t-Test, Left-Tailed

Data: Same as above, test if After < Before.

Step 1: H_0: μ_d = 0, H_a: μ_d > 0 (or d >0?) Actually, Left-tailed → H_a: μ_d <0. Assume negative differences).

If T = −2.5, df=11, α=0.05 → t_{0.05,11} = −1.796
−2.5 < −1.796 → **Reject H_0**

Conclusion: Significant decrease after intervention.

9. Pooled t-Test, Moderate n, Right-Tailed

Data:

- Group 1: $n_1=35$, $\bar{x}_1=52$, $s_1=9$
- Group 2: $n_2=40$, $\bar{x}_2=49$, $s_2=8$

Step 1: Pooled variance

s_p^2 = [(34*81)+(39*64)]/(73) = (2754+2496)/73 = 5250/73 ≈ 71.9

s_p ≈ 8.48

Step 2: SE

SE = 8.48 * sqrt(1/35+1/40) = 8.48 * sqrt(0.02857+0.025) = 8.48 * sqrt(0.05357) ≈ 8.48*0.2316 ≈ 1.964

Step 3: T statistic

T = (52-49)/1.964 = 3/1.964 ≈ 1.528

Step 4: df = 73, t_{0.95} ≈ 1.667

Step 5: Decision

1.528 < 1.667 → **Do not reject H_0**

Conclusion: Insufficient evidence that Group 1 mean is higher.

10. Welch t-Test, Unequal Variances, Two-Sided

Data:

- Group 1: $n_1=10$, $\bar{x}_1=88$, $s_1=5$
- Group 2: $n_2=12$, $\bar{x}_2=82$, $s_2=8$

Step 1: SE

SE = sqrt(25/10 + 64/12) = sqrt(2.5 + 5.333) = sqrt(7.833) ≈ 2.799

Step 2: T

T = (88-82)/2.799 = 6/2.799 ≈ 2.143

Step 3: df

df ≈ $(7.833^2)/((2.5^2/9) + (5.333^2/11))$ = 61.34 / (0.694 + 2.587) = 61.34/3.281 ≈ 18.7 ≈19

Step 4: Critical value (two-sided, α=0.05)

t_{0.975,19} ≈ 2.093

Step 5: Decision

2.143 > 2.093 → **Reject H_0**

Conclusion: Significant difference between means.