Testing the Difference Between Two Means: Large and Small Samples

Dr. Zaid T. Al-Khaledi
Department of Statistics and Informatics
University of Mosul

October 22, 2025

Introduction

In many studies, we are interested in comparing two groups or treatments. For example:

- Do male and female students have the same mean exam score?
- Does a new drug reduce the mean blood pressure more than the old one?
- Are two production lines delivering the same average output?

In such cases, we study the **difference between two population means**:

$$\mu_1 - \mu_2.$$

We test whether this difference is zero (no effect) or significantly different from zero (an effect exists). Depending on the sample size and whether population variances are known or equal, we use one of the following tests:

1. **Large samples:** Z-test for $\mu_1 - \mu_2$.
2. **Small samples:** Two-sample t-test (pooled or Welch version).
3. **Paired samples:** Paired t-test.

Basic Setup

We have two independent samples:

$$X_1, X_2, \ldots, X_{n_1} \quad \text{from population 1, mean } \mu_1, \text{ variance } \sigma_1^2,$$

$$Y_1, Y_2, \ldots, Y_{n_2} \quad \text{from population 2, mean } \mu_2, \text{ variance } \sigma_2^2.$$

The sample means and variances are:

$$\bar{X} = \frac{1}{n_1}\sum X_i, \quad s_1^2 = \frac{1}{n_1 - 1}\sum (X_i - \bar{X})^2,$$

$$\bar{Y} = \frac{1}{n_2}\sum Y_j, \quad s_2^2 = \frac{1}{n_2 - 1}\sum (Y_j - \bar{Y})^2.$$

We are testing:

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \begin{cases} \mu_1 - \mu_2 \neq 0 & \text{(two-sided)} \\ \mu_1 - \mu_2 > 0 & \text{(right-tailed)} \\ \mu_1 - \mu_2 < 0 & \text{(left-tailed).} \end{cases}$$

1 Case 1: Large Samples (Z-Test for $\mu_1 - \mu_2$)

Formulation

When both samples are large ($n_1, n_2 \geq 30$) or population standard deviations σ_1, σ_2 are known, the sampling distribution of $(\bar{X} - \bar{Y})$ is approximately normal:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Under H_0 (i.e. $\mu_1 = \mu_2$),

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$

Decision Rule

Reject H_0 if

$$|Z| > z_{\alpha/2} \quad \text{(two-sided)}, \quad Z > z_\alpha \quad \text{(right-tailed)}, \quad Z < -z_\alpha \quad \text{(left-tailed)}.$$

Example (Hand Computation)

Two factories produce metal rods. Factory A: $n_1 = 50$, $\bar{x} = 10.2$, $\sigma_1 = 0.5$; Factory B: $n_2 = 60$, $\bar{y} = 9.9$, $\sigma_2 = 0.6$. Test at $\alpha = 0.05$ whether their means differ.

$$Z = \frac{10.2 - 9.9}{\sqrt{0.5^2/50 + 0.6^2/60}} = \frac{0.3}{\sqrt{0.005 + 0.006}} = \frac{0.3}{0.103} = 2.91.$$

Critical value $z_{0.025} = 1.96$. Since $|Z| = 2.91 > 1.96$, we **reject** H_0: the factories differ in mean output.

Confidence Interval

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

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Therefore the standard error is:

$$\text{SE} = \sqrt{\frac{0.5^2}{50} + \frac{0.6^2}{60}} = \sqrt{0.005 + 0.006} = \sqrt{0.011} = 0.1049.$$

Margin at $\alpha = 0.05$ ($z_{0.025} = 1.96$):

$$z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1.96 \times 0.1049 = 0.2056$$

Thus the $(1 - \alpha)100\%$ CI for $(\mu_1 - \mu_2)$ is

$$\left(\bar{X} - \bar{Y}\right) \pm z_{\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 0.300 \pm 0.2056 = \boxed{(0.094,\ 0.506)}.$$

2 Case 2: Small Samples (t-Tests)

When σ_1 and σ_2 are unknown and samples are small, we use the t-distribution.

(a) Pooled t-Test (Equal Variances)

Assumptions:

- Both populations are normal.
- Variances are equal ($\sigma_1^2 = \sigma_2^2$).

The pooled variance is:

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

Test statistic:

$$t = \frac{(\bar{X}-\bar{Y}) - (\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{with } df = n_1+n_2-2.$$

Decision rule:

$$|t| > t_{\alpha/2,df} \Rightarrow \text{Reject } H_0.$$

Example. Two diets are tested for weight loss (kg):

Diet A	5.1	4.8	4.9	5.3	5.0	5.1
Diet B	4.4	4.6	4.8	4.3	4.7	4.5

$$\bar{x}_1 = 5.03,\ s_1 = 0.18,\ n_1 = 6, \qquad \bar{x}_2 = 4.55,\ s_2 = 0.18,\ n_2 = 6.$$

$$s_p^2 = \frac{5(0.18^2) + 5(0.18^2)}{10} = 0.0324 \Rightarrow s_p = 0.18.$$

$$t = \frac{5.03 - 4.55}{0.18\sqrt{1/6+1/6}} = \frac{0.48}{0.104} = 4.62.$$

$df = 10$, $t_{0.025,10} = 2.228$. Since $4.62 > 2.228$, we **reject** H_0. The diets differ significantly.

(b) Welch t-Test (Unequal Variances)

If the population variances are not equal, we use Welch's version:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

with approximate degrees of freedom:

$$df \approx \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}.$$

This test does not assume equal variances and is preferred when s_1^2 and s_2^2 differ substantially.

Example. Two teaching methods are compared. Group 1 ($n_1 = 10$, $\bar{x}_1 = 82$, $s_1 = 5.2$), Group 2 ($n_2 = 12$, $\bar{x}_2 = 77$, $s_2 = 8.1$).

$$t = \frac{82-77}{\sqrt{5.2^2/10 + 8.1^2/12}} = \frac{5}{\sqrt{2.704+5.468}} = \frac{5}{2.93} = 1.71.$$

Step 1: Degrees of freedom (Welch–Satterthwaite formula)

$$df = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}} = \frac{\left(5.2^2/10+8.1^2/12\right)^2}{\frac{(2.704)^2}{9}+\frac{(5.468)^2}{11}} = \frac{66.78}{0.812+2.718} = \frac{66.78}{3.530} = 18.9 \approx 19.$$

Thus $df \approx 19$ (rounded).

Step 2: Decision rule. For $\alpha = 0.05$, two-sided, $t_{0.025,19} = 2.093$. Since $|t| = 1.71 < 2.09$, we **fail to reject** H_0.

Confidence Interval (t-based)

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2,df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Standard error:

$$SE = \sqrt{\frac{5.2^2}{10} + \frac{8.1^2}{12}} = 2.93.$$

Margin of error:

$$ME = t_{0.025,19} \times SE = 2.093 \times 2.93 = 6.13.$$

Therefore the 95% CI for $(\mu_1 - \mu_2)$ is

$$(82 - 77) \pm 6.13 = 5.00 \pm 6.13 = \boxed{(-1.13,\ 11.13)}.$$

Interpretation: Because the CI includes 0, there is no significant difference between the two teaching methods at the 5% level.