Testing the Difference Between Two Means: Large and Small Samples

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

November, 03, 2025

1 Case 3: Paired t-Test

When the same subjects are measured twice (before/after treatment, left/right eye, etc.), we analyze the **differences**:

$$d_i = X_i - Y_i.$$

Then test whether the mean difference $\mu_d = 0$:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, \quad df = n - 1.$$

Example (Paired). Blood pressure before and after a drug (mmHg):

$$\text{Before: } 150, 148, 160, 152, 155, \quad \text{After: } 142, 144, 149, 145, 147.$$

Differences: $8, 4, 11, 7, 8$.

$$\bar{d} = 7.6, \ s_d = 2.7, \ n = 5.$$

$$t = \frac{7.6}{2.7/\sqrt{5}} = \frac{7.6}{1.21} = 6.28.$$

Critical $t_{0.025,4} = 2.776$. Since $6.28 > 2.776$, we **reject** H_0: the drug significantly reduces blood pressure.

Interpretation

The paired t-test removes between-subject variability. It is more powerful when the same subjects are measured under both conditions.

One- vs Two-Tailed Tests

- Use **two-tailed** tests when any difference (increase or decrease) matters.
- Use **one-tailed** tests when only one direction is of interest and that direction is specified *before* data collection.

Assumptions Recap

1. Random, independent samples.
2. Normality of each population (or large n by CLT).
3. For pooled t-test: equal variances.
4. For paired t-test: differences are approximately normal.

Summary of Test Selection

Condition	**Test**	**Statistic**
σ_1, σ_2 known or large n	Two-sample Z-test	$Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$
Small n, equal variances	Pooled t-test	$t = \frac{(\bar{X}-\bar{Y})}{s_p\sqrt{1/n_1+1/n_2}}$
Small n, unequal variances	Welch t-test	$t = \frac{(\bar{X}-\bar{Y})}{\sqrt{s_1^2/n_1+s_2^2/n_2}}$
Paired data	Paired t-test	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$

Homework: Two-Sample and Paired Tests

Use $\alpha = 0.05$ for all problems. State hypotheses, choose the correct test (Z, pooled t, Welch t, or paired t), compute the test statistic and critical value(s), and interpret in context.

Q1 (Manufacturing)

Two production lines produce metal screws. From Line A ($n_1 = 40$), $\bar{x}_1 = 50.8\,\text{mm}$, $s_1 = 0.6\,\text{mm}$; from Line B ($n_2 = 35$), $\bar{x}_2 = 51.0\,\text{mm}$, $s_2 = 0.5\,\text{mm}$. Is there evidence of a difference in mean screw length?

Q2 (Education – Unequal Variances, Welch t-Test)

An educator compares two teaching methods. Group 1 ($n_1 = 15$): $\bar{x}_1 = 78$, $s_1 = 8$; Group 2 ($n_2 = 15$): $\bar{x}_2 = 72$, $s_2 = 9$. Assume unequal variances. Is Method 1 more effective?

Q3 (Medical – Paired Data)

A new medication is tested on $n = 10$ patients. For each patient, systolic blood pressure (mmHg) is recorded **before** and **after** treatment.

Patient	1	2	3	4	5	6	7	8	9	10
Before	150	148	155	160	152	149	151	157	153	158
After	143	141	149	151	145	144	146	150	147	149

Use a *paired t-test* at $\alpha = 0.05$ to evaluate whether the drug reduces average systolic blood pressure.

Q4 (Agriculture – Welch t-Test, Unequal Variances)

Two fertilizers are tested on wheat yield (tons/hectare):
Fertilizer A ($n_1 = 12$): $\bar{x}_1 = 3.8$, $s_1 = 0.50$ Fertilizer B ($n_2 = 10$): $\bar{x}_2 = 3.2$, $s_2 = 0.80$
Assume unequal variances and test at $\alpha = 0.05$ whether Fertilizer A produces a higher mean yield. *(Hint: one-sided Welch t-test, $H_1 : \mu_1 > \mu_2$.)*

Candidate Critical Values

Label	Value	Label	Value	Label	Value
$z_{0.10}$	1.282	$z_{0.05}$	1.645	$z_{0.025}$	1.960
$z_{0.01}$	2.326	$z_{0.005}$	2.576	$t_{0.05,\,8}$	1.860
$t_{0.025,\,8}$	2.306	$t_{0.05,\,10}$	1.812	$t_{0.025,\,10}$	2.228
$t_{0.05,\,15}$	1.753	$t_{0.025,\,15}$	2.131	$t_{0.05,\,20}$	1.725
$t_{0.025,\,20}$	2.086	$t_{0.05,\,30}$	1.697	$t_{0.025,\,30}$	2.042
$t_{0.05,\,40}$	1.684	$t_{0.025,\,40}$	2.021	$t_{0.05,\,60}$	1.671
$t_{0.025,\,60}$	2.000	$t_{0.05,\,80}$	1.664	$t_{0.025,\,80}$	1.990

All entries are upper-tail critical values. For two-sided tests use $\alpha/2$ (e.g., $z_{0.025}$ for $\alpha = 0.05$). Match t degrees of freedom to $n - 1$ (one-sample or paired) or to the test's df (pooled or Welch).