Two-Sample Hypothesis Tests for the Difference of Means

AP Statistics · Mr. Merrick · February 22, 2026

We compare two population means by testing a claim about the parameter $\mu_1 - \mu_2$.
From two independent random samples (or two randomized groups), we compute:

$$\overline{x}_1,\ s_1,\ n_1 \qquad \text{and} \qquad \overline{x}_2,\ s_2,\ n_2.$$

For most AP problems, the null value is

$$H_0 : \mu_1 - \mu_2 = 0.$$

1) Two-Sample z test

When to use: σ_1, σ_2 known (rare).

Check conditions:

- Random: each sample is from a random sample or randomized experiment.
- Independence:
  - within groups: $n_1 \le 0.10N_1$ and $n_2 \le 0.10N_2$ (if sampling w/o replacement)
  - between groups: the two samples/groups are independent
- Normal/Large Sample: each population is Normal, or each n is large enough for CLT (≥ 30).

Test statistic (for $H_0 : \mu_1 - \mu_2 = \Delta_0$):

$$\boxed{Z = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

Under H_0, $Z \sim N(0, 1)$.

2) Two-Sample t test (Welch's)

When to use: σ_1, σ_2 unknown (typical).

Check conditions:

- Random: each sample is random or groups are randomized.
- Independence:
  - within groups: $n_1 \le 0.10N_1$ and $n_2 \le 0.10N_2$ (if sampling w/o replacement)
  - between groups: independent samples (or randomized groups)
- Normal/Large Sample:
  - if n_1 and/or n_2 are small: check each group's sample distribution is roughly symmetric with no outliers
  - if both are large (≥ 30): CLT supports the procedure

Standard error (estimated): $SE = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

Test statistic (for $H_0 : \mu_1 - \mu_2 = \Delta_0$):

$$\boxed{T = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}$$

with Welch df:

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

$$\boxed{\min\{n_1 - 1, n_2 - 1\} \le df \le n_1 + n_2 - 2}$$

3) Enrichment: Two-Sample t test with pooled variance (NOT required for AP)

Big idea: If the population variances are equal ($\sigma_1^2 = \sigma_2^2$), we can pool information from both groups to estimate the common variance.

Important: This is *not* needed for AP Statistics. The AP standard is Welch's two-sample t test.

In more advanced settings, equality of variances might be assessed by:

- comparing sample spreads (s_1 vs. s_2), or sample variances (s_1^2 vs. s_2^2),
- using a formal procedure such as Levene's test (beyond AP).

Pooled standard deviation:

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

Pooled standard error:

$$SE_p = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Pooled test statistic (for $H_0 : \mu_1 - \mu_2 = \Delta_0$):

$$\boxed{T = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with} \quad df = n_1 + n_2 - 2}$$

For AP Stats: Use Welch's approximation unless the problem explicitly explores pooling (possible on an investigative task).

Example 1

A manufacturer compares the fill amounts of two machines. From historical calibration, the population standard deviations are known:

$$\sigma_1 = 1.2 \text{ mL}, \qquad \sigma_2 = 1.6 \text{ mL}.$$

A random sample of $n_1 = 40$ bottles from Machine 1 has mean $\overline{x}_1 = 502.3$ mL. A random sample of $n_2 = 35$ bottles from Machine 2 has mean $\overline{x}_2 = 500.9$ mL.
Test, at the $\alpha = 0.05$ level, whether Machine 1 fills *more* on average than Machine 2.

Example 2

A school compares weekly study time for students in two different programs. Two independent random samples are taken.

Group	n	$\overline{x}$ (hours)	s (hours)
Program A	18	6.8	1.9
Program B	14	5.4	2.3

Test, at the $\alpha = 0.10$ level, whether the true mean weekly study time differs between the two programs.

Example 3

A nutritionist compares sodium content (mg) for two brands of soup. Independent samples are taken.

Brand	n	$\overline{x}$ (mg)	s (mg)
Brand 1	10	710	48
Brand 2	9	742	55

Test, at the $\alpha = 0.05$ level, whether Brand 1 has a *lower* true mean sodium content than Brand 2.

Example 4 (Enrichment: pooled t test)

(Not needed for AP.) A researcher believes two populations have equal variances. Independent samples produce:

Group	n	$\overline{x}$	s
Group 1	22	15.2	3.1
Group 2	20	12.9	2.9

Test, at the $\alpha = 0.05$ level, whether Group 1 has a higher true mean than Group 2.