TWO-SAMPLE z TESTS FOR $p_1 - p_2$

AP Statistics · Mr. Merrick · February 9, 2026

We often want to compare two groups to determine whether there is convincing statistical evidence that their population proportions differ.

$$p_1 - p_2 = (\text{true proportion in Group 1}) - (\text{true proportion in Group 2})$$

A hypothesis test evaluates whether an observed difference is likely due to random chance or reflects a real difference in the populations.

A two sample z-test for $p_1 - p_2$ is used to test:

$$H_0: \ p_1 - p_2 = 0 \qquad \text{vs} \qquad H_a: \ p_1 - p_2 \neq 0, \ < 0, \ \text{or} \ > 0$$

Test statistic

Under H_0, we assume $p_1 = p_2$, so we use a **pooled proportion**:

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

The test statistic is:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

Conditions

This test relies on the sampling distribution of $\hat{p}_1 - \hat{p}_2$ (assuming H_0 is true) being approximately Normal.

- **Random:** both samples come from random sampling or random assignment.

- **Independence:** each sample is less than 10% of its population.

- **Large Counts (pooled):**

$$n_1\hat{p} \geq 10, \quad n_1(1 - \hat{p}) \geq 10, \quad n_2\hat{p} \geq 10, \quad n_2(1 - \hat{p}) \geq 10$$

If these conditions are met, the two-sample z test provides reliable results.

Example: Online checkout completion

An online retailer is comparing two website designs to see whether they affect checkout completion rates.

- Design A: $n_1 = 500$ users, $x_1 = 365$ completed checkout.

- Design B: $n_2 = 480$ users, $x_2 = 330$ completed checkout.

Test at the $\alpha = 0.05$ level whether the checkout completion rates differ between the two designs.

Practice

Helmet use

A city surveys cyclists in two neighborhoods to compare helmet use.

- Neighborhood A: $n_1 = 300$, $x_1 = 198$ wear helmets.

- Neighborhood B: $n_2 = 260$, $x_2 = 143$ wear helmets.

Test at the $\alpha = 0.05$ level whether helmet use differs between the neighborhoods.

Customer satisfaction

A company compares satisfaction rates for customers using two support systems.

- Chat support: $n_1 = 420$, $x_1 = 336$ satisfied.

- Phone support: $n_2 = 390$, $x_2 = 285$ satisfied.

Test at the $\alpha = 0.10$ level whether chat support has a higher satisfaction rate.

What does a small p-value mean?