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.

Step 1 (Parameter + hypotheses).
Let $p_1 - p_2$ be the true difference in checkout completion proportions for Design A and Design B (A minus B).
$\alpha = 0.05$

$$H_0: \ p_1 - p_2 = 0 \qquad H_a: \ p_1 - p_2 \neq 0$$

Step 2 (Conditions).
Random: users were randomly assigned to one of the two website designs.
Independence: each group represents less than 10% of all site users.
Pooled proportion:

$$\hat{p} = \frac{365 + 330}{500 + 480} = \frac{695}{980} \approx 0.709$$

Large Counts:

$$500(0.709),\ 500(0.291),\ 480(0.709),\ 480(0.291) \geq 10$$

Since conditions are met, a two-sample z test is appropriate.

Step 3 (Test statistic + p-value).

$$\hat{p}_1 = \frac{365}{500} = 0.730, \qquad \hat{p}_2 = \frac{330}{480} = 0.688$$

$$SE = \sqrt{\frac{0.709(0.291)}{500} + \frac{0.709(0.291)}{480}} \approx 0.029$$

$$z = \frac{(0.730 - 0.688) - 0}{0.029} \approx 1.46$$

Two-sided p-value:

$$p\text{-value} \approx 0.143$$

Step 4 (Decision + conclusion).
Since p-value > 0.05, we fail to reject H_0.
There is insufficient evidence at the 5% level to conclude that checkout completion rates differ between the two website 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.

Step 1 (Parameter + hypotheses).
Let $p_1 - p_2$ be the true difference in helmet use proportions (Neighborhood A minus Neighborhood B).
$\alpha = 0.05$

$$H_0: \ p_1 - p_2 = 0 \qquad H_a: \ p_1 - p_2 \neq 0$$

Step 2 (Conditions).
Random: cyclists were randomly selected in each neighborhood.
Independence: each sample is less than 10% of the cyclist population.
Pooled proportion:

$$\hat{p} = \frac{198 + 143}{300 + 260} = \frac{341}{560} \approx 0.609$$

Large Counts satisfied.

Step 3 (Test statistic + p-value).

$$\hat{p}_1 = 0.660, \qquad \hat{p}_2 = 0.550$$

$$SE \approx 0.041, \quad z \approx 2.66$$

Two-sided p-value ≈ 0.0078.

Step 4 (Decision + conclusion).
Reject H_0. There is sufficient evidence at the 5% level to conclude that helmet use differs between the two 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.

Step 1 (Parameter + hypotheses).
Let $p_1 - p_2$ be the true difference in satisfaction proportions (chat minus phone).
$\alpha = 0.10$

$$H_0: \ p_1 - p_2 = 0 \qquad H_a: \ p_1 - p_2 > 0$$

Step 2 (Conditions).
Random: customers were randomly sampled.
Independence: each sample is less than 10% of customers.
Pooled proportion:

$$\hat{p} = \frac{336 + 285}{420 + 390} = \frac{621}{810} \approx 0.767$$

Large Counts satisfied.

Step 3 (Test statistic + p-value).

$$\hat{p}_1 = 0.800, \qquad \hat{p}_2 = 0.731$$

$$SE \approx 0.030, \quad z \approx 2.33$$

Right-tailed p-value ≈ 0.010.

Step 4 (Decision + conclusion).
Reject H_0. There is sufficient evidence at the 10% level to conclude that chat support has a higher satisfaction rate.

What does a small p-value mean?

A small p-value means that if the null hypothesis were true, observing a sample difference at least as extreme as the one observed would be unlikely due to random chance alone.