# Two-Sample $z$ Intervals for $p_1 - p_2$

*AP Statistics · Mr. Merrick · February 9, 2026*

We often want to compare two groups by estimating a difference in population proportions.

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

A confidence interval gives a range of plausible values for $p_1 - p_2$.

A two sample $z$-interval for $p_1 - p_2$ is given by:

$$(\hat{p}_1 - \hat{p}_2) \ \pm \ z^\star \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- $\hat{p}_1 = \dfrac{x_1}{n_1}, \quad \hat{p}_2 = \dfrac{x_2}{n_2}$
- $z^\star = 1.645$ (90%), 1.96 (95%), 2.576 (99%)

### Conditions

This method relies on the sampling distribution of the statistic $\hat{p}_1 - \hat{p}_2$ being approximately Normal and having a predictable standard deviation. The following conditions ensure that these assumptions are reasonable.

- **Random:** Random sampling or random assignment helps ensure that the samples are representative of their populations. Without randomness, the results may be biased, and the confidence interval may not reflect the true population difference.
- **Independence:** When sampling without replacement, the 10% condition ensures that the outcome of one observation does not meaningfully affect another. This allows us to treat observations as independent and use the standard error formula in the interval.
- **Large Counts (both groups):** Having at least 10 successes and 10 failures in each group ensures that the sampling distributions of $\hat{p}_1$ and $\hat{p}_2$ are approximately Normal. This, in turn, makes the distribution of their difference approximately Normal, which is required for the $z$ interval.

If these conditions are satisfied, the two-sample $z$ interval provides reliable results.

## Example: Seatbelt use

A state compares seatbelt use between two regions.

- Urban region: $n_1 = 200$, $x_1 = 162$ wear seatbelts.
- Rural region: $n_2 = 180$, $x_2 = 126$ wear seatbelts.

Construct a 95% confidence interval for $p_1 - p_2$ (urban minus rural).

# Practice

## Voter turnout

A researcher compares voter turnout in two age groups.

- Ages 18–29: $n_1 = 250$, $x_1 = 140$ voted.
- Ages 30–49: $n_2 = 220$, $x_2 = 154$ voted.

Construct and interpret a 95% confidence interval for $p_1 - p_2$ (18–29 minus 30–49).

### Product defects

Two factories produce the same part.

- Factory A: $n_1 = 400$, $x_1 = 28$ defective.
- Factory B: $n_2 = 350$, $x_2 = 42$ defective.

Construct and interpret a 90% confidence interval for $p_1 - p_2$ (A minus B).

### What does it mean if 0 is inside the interval?