The one-sample z interval for a proportion

A random sample found that some share of the people asked have a feature. What range of values for the share in the whole population does it support, and how many people must you ask to pin the share down to a chosen margin?

WORDS TO KNOW

p: the parameter, the share of the whole population with the feature. It is a fixed number you do not know.

z^* (z star): the critical value for the confidence level: 1.645, 1.96 and 2.576 for 90%, 95% and 99%.

p^* (p star): a planning guess for p, used to choose a sample size before any data exist.

$\hat{p}$ (p-hat): the sample proportion, $\hat{p} = x \div n$, with x successes in a sample of n.

Margin of error: the amount added to and subtracted from $\hat{p}$.

KEY IDEAS

1. The interval, in four steps

The interval is "estimate, plus or minus a critical value times a standard error." The standard error uses $\hat{p}$ in place of the unknown p.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Parameter. Name p: the share of a stated population with a stated feature.

Conditions. Random; n at most a tenth of the population; at least 10 successes and 10 failures.

Compute. $\hat{p}$, the standard error, the margin, the two ends.

Conclude. "We are 95% confident that between ... and ... of all ... have ..."

The level belongs to the method: about 95% of all possible random samples give an interval that captures p. A calculator's 1-PropZInt does the arithmetic, but it checks no condition.

2. What the interval rules in and out

Values inside the interval are plausible values of p at that level; values outside are not. A plausible value is not proven: every other value inside is plausible too.

3. Choosing the sample size

Before a survey, ask for a margin no larger than m and solve for n.

$$n \geq \left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

With no prior estimate, use the safe guess $p^* = 0.5$: it makes $p^*(1-p^*)$ as large as it can be, 0.25, so the margin holds whatever the share is. Always round n up.

Free sample from math.wiki for classroom use. Do not post. math.wiki

WATCH OUT

- **"95% confident" is not a 95% chance for this interval.** The parameter p is fixed: this interval either contains it or does not. The 95% describes the method.
- **The interval is about the population, not the sample.** The share in the sample is already known exactly.
- **Round the sample size up, even for a small decimal.** Rounding down leaves the margin just above the target.
- **A guess other than 0.5 can fall short.** If the sample share lands nearer 0.5 than p^*, the margin comes out larger than planned.

WORKED EXAMPLES

Example 1. A university has about 48,000 students. A simple random sample of 320 finds 104 who ride a bike to campus. Construct and interpret a 95% confidence interval.

1. *Parameter.* Let p be the proportion of all the university's students who ride a bike to campus.
2. *Conditions.* A simple random sample. A tenth of 48,000 is 4,800, and 320 is far below it. There are 104 successes and 320 − 104 = 216 failures, both at least 10.
3. *Compute.* With $z^* = 1.96$:

$$\hat{p} = \frac{104}{320} = 0.325 \qquad \text{SE} = \sqrt{\frac{0.325 \times 0.675}{320}} \approx 0.0262 \qquad 1.96 \times 0.02618 \approx 0.0513$$

   The interval is 0.325 ± 0.0513, from 0.2737 to 0.3763.
4. *Conclude.* We are 95% confident that between 27.4% and 37.6% of all the university's students ride a bike to campus.

Answer: about 0.274 to 0.376. **Check:** the center $(0.2737 + 0.3763) \div 2$ is 0.325, which is $\hat{p}$.

Example 2. Using the interval from Example 1, a campus paper says "one in four" students bike, and a planner says 35%. Which claims are plausible?

1. One in four is 0.25, below the lower end 0.274. Not plausible at 95%.
2. 0.35 lies inside the interval, so it is plausible. That does not prove it: 0.30 is just as plausible.

Answer: at 95% confidence, the sample is evidence that more than one in four students bike; 35% fits the data, but so does every value from 0.274 to 0.376.

Example 3. A town wants a 90% interval with a margin of error of at most 0.04. How many people should it sample with no prior estimate? With a pilot estimate of 0.25?

1. With $z^* = 1.645$: $(1.645 \div 0.04)^2 = 41.125^2 = 1{,}691.265625$.
2. Safe guess: 1,691.27 × 0.25 ≈ 422.82. Round up to 423.
3. Pilot guess: $p^*(1 - p^*) = 0.25 \times 0.75 = 0.1875$, and 1,691.27 × 0.1875 ≈ 317.11. Round up to 318.

Answer: 423 people with no estimate, or 318 with the guess 0.25. The smaller sample is planned for a share of 0.25: if the sample share lands nearer 0.5, the margin can come out above 0.04.

Free sample from math.wiki for classroom use. Do not post. math.wiki