Inference for Categorical Data: Proportions

AP Statistics

Why Be Normal?

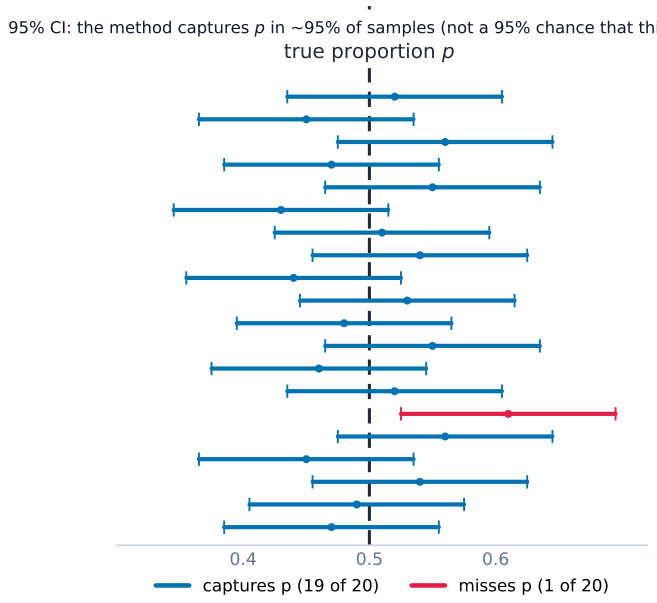
Because a sample proportion $\hat{p}$ is approximately **normally** distributed (when the conditions hold), we can measure how far a sample result is from a claimed value in standard errors, and turn that into a probability. This is what makes **inference** 推断 – drawing conclusions about a population from a sample – possible.

Confidence Interval for a Proportion

A **confidence interval** 置信区间 estimates the parameter as a range: statistic ± **margin of error** 误差幅度.

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

z^* is the critical value for the **confidence level** 置信水平 (e.g. 1.96 for 95%). Conditions: **random** sample, **Large Counts** ($n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$), and the **10%** condition. Interpret it: "We are 95% confident the true proportion of... is between... and...". Interpret the **level**: "In 95% of samples, this method produces an interval that captures the true proportion."

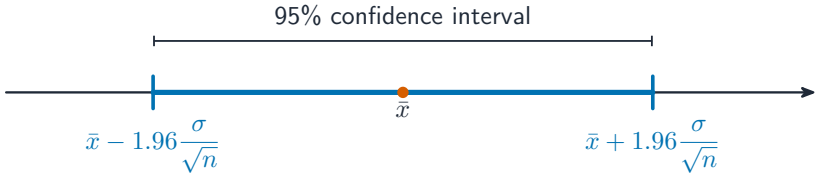


Over many samples, about 95% of 95% confidence intervals capture the true proportion

Worked example. In a random sample of 200 people, 120 support a policy, so $\hat{p} = 0.60$. A 95% interval uses $z^* = 1.96$:

$$0.60 \pm 1.96\sqrt{\frac{0.60(0.40)}{200}} = 0.60 \pm 0.068 = (0.532,\ 0.668).$$

We are 95% confident the true proportion of supporters is between 53.2% and 66.8%.



A 95% confidence interval reaches 1.96 standard errors each side of the estimate

Choosing the sample size. To keep the margin of error no larger than a target m, set $z^*\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \le m$ and solve for n. When you have no estimate of $\hat{p}$, use $\hat{p} = 0.5$: it makes $\hat{p}(1-\hat{p})$ as large as possible, giving the **safe (largest) required sample size**. Always round the result **up** to the next whole person.

Worked example. How many people must you survey for a 95% interval with margin of error at most 0.03? Using $\hat{p} = 0.5$ and $z^* = 1.96$:

$$n = \frac{(z^*)^2\,\hat{p}(1-\hat{p})}{m^2} = \frac{1.96^2(0.5)(0.5)}{0.03^2} = \frac{0.9604}{0.0009} \approx 1067.1,$$

so you survey 1068 people (always round **up**, since 1067 would leave the margin a shade too big).

Justifying a Claim from an Interval

To judge a claimed value: if it lies **inside** the interval, the data are consistent with it; if it lies **outside**, the data give evidence against it. Base the conclusion on whether the plausible values include the claim, in context.

Setting Up a Test for a Proportion

A **significance test** 显著性检验 weighs evidence against a claim. State a **null hypothesis** 原假设 H_0 and an **alternative hypothesis** 备择假设 H_a about the parameter p:

$$H_0 : p = p_0 \qquad H_a : p \neq p_0 \text{ (or } <, >).$$

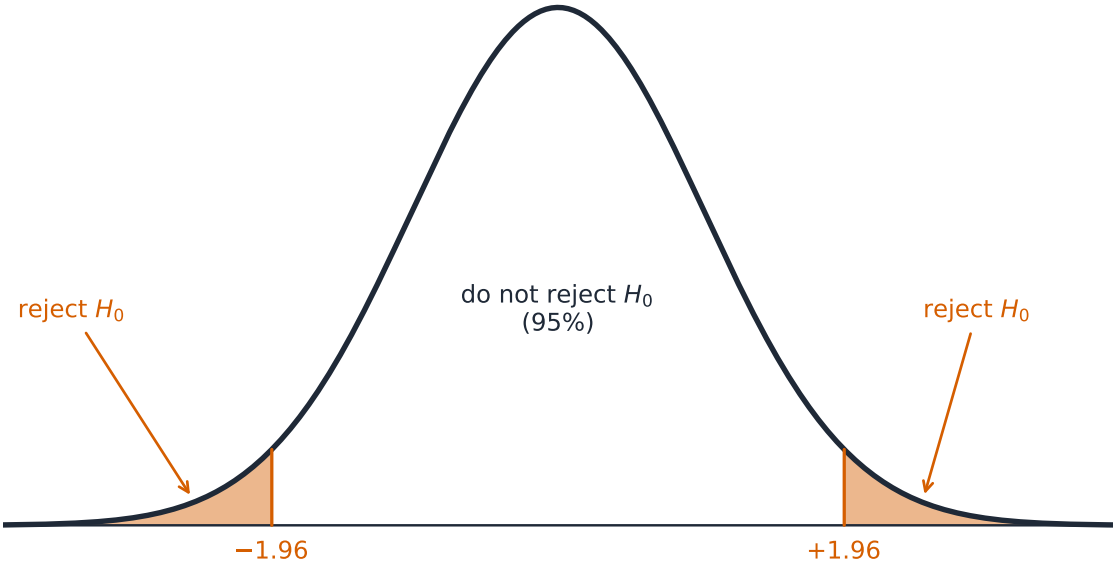
Check the same conditions (random, Large Counts using p_0, 10%). The **test statistic** 检验统计量 counts standard errors from p_0:

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}.$$

Worked example. A company claims 90% satisfaction ($p_0 = 0.90$); a sample of 100 finds 84 satisfied ($\hat{p} = 0.84$). Test $H_0 : p = 0.90$ vs $H_a : p \neq 0.90$ at $\alpha = 0.05$:

$$z = \frac{0.84 - 0.90}{\sqrt{0.90(0.10)/100}} = \frac{-0.06}{0.03} = -2.0,$$

giving a two-tailed p-value of about $2(0.023) = 0.046$. Since $0.046 < 0.05$, **reject** H_0 – there is evidence the true satisfaction rate differs from (is below) 90%.



A two-tailed 5% test rejects the null hypothesis in the shaded tails

Interpreting p-Values

The *p*-**value** P 值 is the probability of getting a sample result **as extreme or more extreme** than the observed one, **assuming** H_0 **is true**. A small *p*-value means the data would be surprising if H_0 held – evidence against H_0. It is *not* the probability that H_0 is true.

Concluding a Test

Compare the *p*-value to the **significance level** 显著性水平 α (often 0.05):

- $p \leq \alpha$: **reject** H_0 – there is convincing evidence for H_a.
- $p > \alpha$: **fail to reject** H_0 – not enough evidence for H_a (never "accept H_0").

Always write the conclusion **in context**, linking back to the claim.

Type I and Type II Errors

- A **Type I error** 第一类错误: rejecting a **true** H_0 (a false alarm). Its probability is α.
- A **Type II error** 第二类错误: failing to reject a **false** H_0 (a missed detection). Its probability is β.
- The **power** 检验效能 $= 1 - \beta$ is the chance of correctly detecting a real effect. Power rises with a larger sample, a larger effect, or a larger α.

Describe each error and its consequence in the problem's context.

Confidence Interval for a Difference of Proportions

To compare two proportions, estimate $p_1 - p_2$:

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

Conditions must hold in **both** samples, and the samples must be independent.

Justifying a Claim About Two Proportions

If the interval for $p_1 - p_2$ contains 0, the data are consistent with **no difference**; if it lies entirely above or below 0, there is evidence of a difference (in that direction). State the direction and context.

Setting Up a Test for a Difference

Hypotheses compare the two proportions: $H_0 : p_1 = p_2$ versus $H_a : p_1 \neq p_2$ (or $<, >$). Because H_0 says the proportions are equal, use a **combined (pooled)** 合并 sample proportion $\hat{p}_c = \dfrac{\text{total successes}}{\text{total sample size}}$ to estimate the common p.

Carrying Out a Test for a Difference

The pooled two-proportion z statistic:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Find the p-value from the normal model, compare to α, and conclude in context – the same four-step logic as the one-proportion test.

Exam tips

- State the conditions (random, 10%, large counts np, $nq \geq 10$) before any proportion inference.
- A **confidence interval** = estimate $\pm$ margin of error; "95% confident" refers to the **method's** long-run capture rate.
- For a **test**, write H_0 and H_a, compute the test statistic, find the p-value, and compare to α.
- A small p-value is evidence **against** H_0; failing to reject does **not** prove H_0.
- Larger samples shrink the margin of error; a higher confidence level widens it.