PH 702 Introduction to Biostatistics

Part 8: Inference for Population Proportions

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2026/11/16 23:47:49

Learning objectives

- Distinguish a population proportion p from a sample proportion $\hat{p}$.
- Construct and interpret confidence intervals for one proportion OR a difference of proportions.
- Test hypotheses about one or two population proportions using P-values.
- Check the assumptions of each method and identify when a plus-four interval is useful.
- Plan sample size and interpret relative risk.

Data type 1: One SRS from a population with an UNKNOWN proportion p

- Binary outcome, success and failure.
- p: population proportion of successes (unknown parameter).
- X: number of successes among n sampled observations.
- $\hat{p} = X/n$: sample proportion.
- For sufficiently large success and failure counts, the sampling distribution of $\hat{p}$ is approximately normal.

$(1-\alpha)\times 100\%$ large-sample (Wald) confidence interval for p

$$\hat{p} \pm z_{1-\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- $z_{1-\alpha/2}$: the $(1-\alpha/2)\times 100\%$ quantile of standard normal.
- Margin of error: $z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$.
- For this large-sample interval, we expect both $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$ (at least 10 successes and 10 failures).
- Interpretation: We are $(1-\alpha)\times 100\%$ confident that the interval covers the population proportion p.

Demo: Experts' views on robotics

A Pew survey asked whether robotic devices would displace more jobs than they created by 2025. Of 1,896 experts, 910 answered yes. Construct a 90% confidence interval for the proportion of experts who expected a net decrease in jobs.

```
data robotics_ci;
  x=910; n=1896; alpha=0.10;
  phat=x/n;
  z=quantile('NORMAL',1-alpha/2);
```

```
  se=sqrt(phat*(1-phat)/n);
  lower=phat-z*se; upper=phat+z*se;
  format phat lower upper 8.4;
run;
```

Exercise: Red Smarties

A package contains 122 Smarties, of which 19 are red. Calculate a 95% large-sample confidence interval for the proportion of red Smarties.

Plus-four adjustment

When the sample is relatively small, the large-sample confidence interval may perform poorly. The plus-four adjustment improve the previous confidence interval by adding two hypothetical successes and two hypothetical failures. The resulting $(1-\alpha)\times 100\%$ plus-four confidence interval is

$$\tilde{p} \pm z_{1-\alpha/2} \times \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}.$$

- $\tilde{p} = (X+2)/(n+4)$
- Use this interval when the confidence level is at least 90% and the sample size n is at least 10, with any counts of successes or failures less than 10.

Exercise: Defective items

In an SRS of 20 items, 6 are defective. Should you use the ordinary large-sample Wald interval or the plus-four interval for the population proportion defective?

Minimum sample size for inferring a proportion

To target a specific margin of error, say m, at confidence level $1-\alpha$, we need sample size n such that

$$n \geq \frac{z^2_{1-\alpha/2} \times p^*(1-p^*)}{m^2}$$

That is, n is at least as large as the smallest integer $\geq z^2_{1-\alpha/2} \cdot p^*(1-p^*)/m^2$.

- p^* is a guess on the proportion from prior information. If no reliable estimate exists, use $p^* = 0.5$ for a conservative sample size (because $(.5)(1-.5) \geq p^*(1-p^*)$ for all $p^* \in (0,1)$).

Demo: Campus satisfaction

A university wants to estimate the proportion of students satisfied with campus life, requiring the margin of error to be as low as 0.03 at 95% confidence. Without prior knowledge, use $p^* = 0.5$.

$$n \geq \frac{1.96^2(0.5)(0.5)}{0.03^2} = 1067.11\ldots$$

```
data campus_size;
  alpha=0.05; m=0.03; pstar=0.5;
  z=quantile('NORMAL',1-alpha/2);
  n_min=ceil(z**2*pstar*(1-pstar)/(m**2));
run;
```

One-sample z test of (significance) level α for hypotheses $H_0 : p = p_0$ vesus $H_a : p > p_0$, $H_a : p < p_0$, or $H_a : p \neq p_0$

The test statistic is

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

- Note that we use p_0, not $\hat{p}$, in the denominator.
- Expect both $np_0 \geq 10$ and $n(1-p_0) \geq 10$.
- P-value: $\Pr(Z \geq z_{\text{obs}})$ for $H_a : p > p_0$; $\Pr(Z \leq z_{\text{obs}})$ for $H_a : p < p_0$; $2\Pr(Z \geq |z_{\text{obs}}|)$ for $H_a : p \neq p_0$, where Z follows the standard normal.
- Reject H_0 when the P-value $\leq \alpha$; otherwise fail to reject H_0.
- Demo conclusion: Testing hypotheses H_0 : ___ vs. H_a : ___ , we carried out the ___ test and obtained P-value ___ . So, at level ___ , there was/wasn't sufficient statistical evidence against H_0, i.e., we believed that ___ .

Demo: Sunblock protection

A testing company compared a sunblock product with its competitor's product on 20 randomly chosen subjects. The company's product performed better for 13 subjects. Is there evidence that it offers superior protection?

$$H_0 : p = 0.50 \qquad \text{vs.} \qquad H_a : p > 0.50.$$

```
data sunblock_test;
  x=13; n=20; p0=0.50; alpha=0.05;
  phat=x/n;
  z=(phat-p0)/sqrt(p0*(1-p0)/n);
  pvalue=1-cdf('NORMAL', z, 0, 1);
run;
```

Conclusion: Testing hypotheses H_0 : ___ vs. H_a : ___ , we carried out the ___ test and obtained P-value ___ . So, at level ___ , there was/wasn't sufficient statistical evidence against H_0, i.e., we believed that ___ .

Data type 2: Two independent SRSs from populations with UNKNOWN proportions p_1 and p_2

- X_1 successes in sample 1 of size n_1; $\hat{p}_1 = X_1/n_1$.
- X_2 successes in sample 2 of size n_2; $\hat{p}_2 = X_2/n_2$.

$(1-\alpha) \times 100\%$ large-sample (Wald) confidence interval for $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2} \times \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

- Check adequate observed successes and failures in each group (at least 10 of each).
- If the interval includes/excludes 0, we are $(1-\alpha)\times 100\%$ confident that there isn't/is difference in the two population proportions.

Demo: Instagram use

Of 537 young women surveyed, 328 reported using Instagram; of 532 young men, 234 reported using Instagram. Estimate $p_1 - p_2$ (women minus men) with a 95% confidence interval.

```
data instagram_ci;
  x1=328; n1=537; x2=234; n2=532; alpha=0.05;
  p1=x1/n1; p2=x2/n2; difference=p1-p2;
  se=sqrt(p1*(1-p1)/n1+p2*(1-p2)/n2);
  zstar=quantile('NORMAL',1-alpha/2);
  lower=difference-zstar*se; upper=difference+zstar*se;
run;
```

Plus-four adjustment

Add one success and one failure to each group (four added observations total):

$$(\tilde{p}_1 - \tilde{p}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}.$$

- $\tilde{p}_1 = (X_1+1)/(n_1+2)$ and $\tilde{p}_2 = (X_2+1)/(n_2+2)$
- Use this interval when the confidence level is at least 90% and sample sizes n_1 and n_2 are both ≥ 10, with any counts of successes or failures in any sample less than 10.

Demo: Tanner scores

A pilot study included an SRS of 12 girls and an SRS of 12 boys. Three girls and four boys had Tanner scores of 4 or 5. Construct a 95% plus-four interval for $p_{\text{girls}} - p_{\text{boys}}$, where p_{girls} and p_{boys} denote the girls' and boys' proportions of getting Tanner scores of at least 4, respectively.

```
data tanner_plus4;
  x1=3; n1=12; x2=4; n2=12; alpha=0.05;
  p1=(x1+1)/(n1+2); p2=(x2+1)/(n2+2);
  difference=p1-p2;
  se=sqrt(p1*(1-p1)/(n1+2)+p2*(1-p2)/(n2+2));
  zstar=quantile('NORMAL',1-alpha/2);
  lower=difference-zstar*se; upper=difference+zstar*se;
run;
```

Exercise: Two manufacturing methods

A manufacturer makes 10 prototypes with each of two methods. There are 3 defective items using method 1 and 5 defective items using method 2.

- Why is an ordinary large-sample confidence interval questionable?
- Construct a 95% plus-four confidence interval for $p_1 - p_2$.

- Does the interval provide evidence that the defect rates differ?

Two-sample z test of (significance) level α for hypotheses $H_0 : p_1 - p_2 = 0$ vesus $H_a : p_1 - p_2 < 0$, $H_a : p_1 - p_2 > 0$, or $H_a : p_1 - p_2 \neq 0$

The test statistic is

$$z_{\text{obs}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_{\text{pool}}(1 - \hat{p}_{\text{pool}})(1/n_1 + 1/n_2)}}.$$

- $\hat{p}_{\text{pool}} = (X_1 + X_2)/(n_1 + n_2)$: the pooled estimate of the common proportion under H_0
- Verify independent random samples and adequate expected success/failure counts in both groups under H_0.
- For $H_a : p_1 - p_2 < 0$, P-value $= \Pr(Z \leq z_{\text{obs}})$; for $H_a : p_1 - p_2 > 0$, P-value $= \Pr(Z \geq z_{\text{obs}})$; for $H_a : p_1 - p_2 \neq 0$, P-value $= 2\Pr(Z \geq |z_{\text{obs}}|)$. Z follows the standard normal distribution.
- Demo conclusion: Testing hypotheses H_0 : ___ vs. H_a : ___ , we carried out the ___ test and obtained P-value ___ . So, at level ___ , there was/wasn't sufficient statistical evidence against H_0, i.e., we believed that ___ .

Demo: Instagram use (continued)

Test whether young women and young men are equally likely to report Instagram use, using the same data: women 328/537 and men 234/532.

$$H_0 : p_1 = p_2 \qquad \text{vs.} \qquad H_a : p_1 \neq p_2.$$

```
data instagram_test;
  x1=328; n1=537; x2=234; n2=532; alpha=0.05;
  p1=x1/n1; p2=x2/n2;
  ppool=(x1+x2)/(n1+n2);
  se0=sqrt(ppool*(1-ppool)*(1/n1+1/n2));
  z=(p1-p2)/se0;
  pvalue=2*(1-cdf('NORMAL', abs(z), 0, 1));
run;
```

Conclusion: Testing hypotheses H_0 : ___ vs. H_a : ___ , we carried out the ___ test and obtained P-value ___ . So, at level ___ , there was/wasn't sufficient statistical evidence against H_0, i.e., we believed that ___ .

Exercise: Breakfast habits

In a sample of 446 students, 246 ate breakfast. Test $H_0 : p = 0.50$ against $H_a : p > 0.50$. What is your conclusion?

Exercise: Compare the confidence interval and the test

Using the Instagram example:

- Does the 95% confidence interval for $p_1 - p_2$ include 0?
- Does the two-sided hypothesis test reject $H_0 : p_1 - p_2 = 0$ at $\alpha = 0.05$?
- Do these two conclusions agree?

Exercise: Treatment success

A new treatment is given to 150 patients, with 115 successes. A standard treatment is given to 100 patients, with 65 successes.

- State the null and alternative hypotheses for the claim that the new treatment has a higher success probability.
- State the conclusion at $\alpha = 0.05$.

$(1-\alpha)\times 100\%$ approximate confidence interval for relative risk $RR = p_1/p_2$

The relative risk $RR = p_1/p_2$ compares the event probabilities in the two populations.

- $RR = 1$: equal event probabilities.
- $RR > 1$: the event is more likely in Population 1.
- $RR < 1$: the event is less likely in Population 1.

The $(1-\alpha)\times 100\%$ approximate (Wald) confidence interval for $\ln(RR)$ is

$$\ln(\widehat{RR}) \pm z_{1-\alpha/2}\sqrt{\left(\frac{1}{X_1}-\frac{1}{n_1}\right)+\left(\frac{1}{X_2}-\frac{1}{n_2}\right)}.$$

- $\widehat{RR} = \hat{p}_1/\hat{p}_2$

Exponentiate the lower and upper limits to obtain the confidence interval for RR:

$$\left[\widehat{RR}\Big/ \exp\left\{z_{1-\alpha/2}\sqrt{\left(\frac{1}{X_1}-\frac{1}{n_1}\right)+\left(\frac{1}{X_2}-\frac{1}{n_2}\right)}\right\},\quad \widehat{RR}\times \exp\left\{z_{1-\alpha/2}\sqrt{\left(\frac{1}{X_1}-\frac{1}{n_1}\right)+\left(\frac{1}{X_2}-\frac{1}{n_2}\right)}\right\}\right]$$

Concluding application: A/B testing

A/B testing is a randomized comparative experiment in which participants are randomly assigned to one of two groups:

- A (control): The current version.
- B (treatment): A new version we want to evaluate.

A/B testing is widely used to evaluate changes in websites, healthcare interventions, and public health programs.

Example: Conversion rate

Suppose a website wants to know whether design B increases the percentage of visitors who click a registration button. Assume that each visitor either clicks the button or does not click it. Which statistical method is appropriate?

(a) One-sample z test

(b) Two-sample z test

(c) Two-sample t test