

Section 6.1

Inference for a Single Proportion

Stats 7 Summer Session II 2022

Practice

Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?

- (a) All 1000 get the drug
- (b) 500 get the drug, 500 don't

Results from the General Social Survey (GSS)

The GSS asks the same question, below is the distribution of responses from the 2010 survey:

All 1000 get the drug	99
500 get the drug 500 don't	571
<hr/>	
Total	670

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer “500 get the drug 500 don't”? What are the parameter of interest and the point estimate?

Parameter of interest: proportion of *all* Americans who have good intuition about experimental design.

p a population proportion

Point estimate: proportion of *sampled* Americans who have good intuition about experimental design.

\hat{p} a sample proportion

Inference on a proportion

What percent of all Americans have good intuition about experimental design, i.e. would answer "500 get the drug 500 don't"?

We can answer this research question using a confidence interval, which we know is always of the form

$$\text{point estimate} \pm ME$$

And we also know that $ME = \text{critical value times the } SE$ of the point estimate.

$$SE_{\hat{p}} = ?$$

Standard error of a sample proportion

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Sample proportions are also nearly normally distributed

Central limit theorem for proportions:

Sample proportions will be nearly normally distributed with mean equal to the population mean, p , and standard error equal to

$$\hat{p} \sim N \left(\text{mean} = p, SE = \sqrt{\frac{p(1-p)}{n}} \right)$$

- But of course this is true only under certain conditions...

independent observations, at least 10 successes and 10 failures

Note: If p is unknown (most cases), we use \hat{p} in the calculation of the standard error.

Back to experimental design...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design?

Given: $n = 670$, $\hat{p} = 0.85$. First check conditions.

1. *Independence*: The sample is random, and $670 < 10\%$ of all Americans, therefore we can assume that one respondent's response is independent of another.
2. *Success-failure*: 571 people answered correctly (successes) and 99 answered incorrectly (failures), both are greater than 10.

Practice

We are given that $n = 670$, $\hat{p} = 0.85$, we also just learned that the standard error of the sample proportion is

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Which of the below is the correct calculation of the 95% confidence interval?

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}}$ (0.82, 0.88)

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%.

$$ME = z^* \times SE$$

$$0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \text{Use estimate for } \hat{p} \text{ from previous study}$$

$$0.01^2 \geq 1.96^2 \times \frac{0.85 \times 0.15}{n}$$

$$n \geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2}$$

$$n \geq 4898.04 \rightarrow n \text{ should be at least 4,899}$$

What if there isn't a previous study?

... use $\hat{p} = 0.5$

why?

- if you don't know any better, 50-50 is a good guess
- $\hat{p} = 0.5$ gives the most conservative estimate -- highest possible sample size

CI vs. HT for proportions

Success-failure condition:

- CI: At least 10 *observed* successes and failures
- HT: At least 10 *expected* successes and failures, calculated using the null value

Standard error:

- CI: calculate using observed sample proportion:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

- HT: calculate using the null value:

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$

Practice

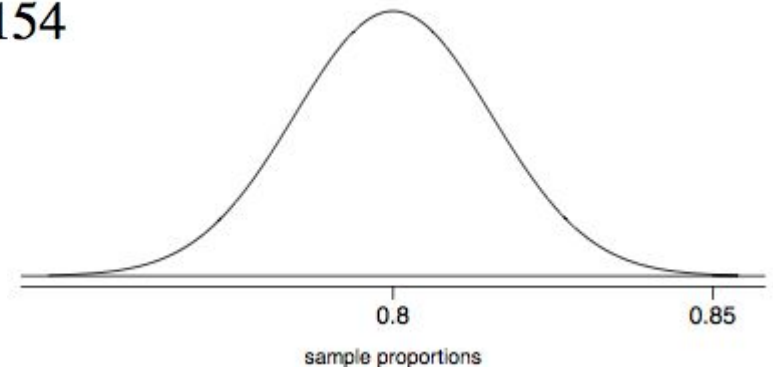
The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \quad H_A : p > 0.80$$

$$SE = \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154$$

$$Z = \frac{0.85 - 0.80}{0.0154} = 3.25$$

$$p\text{-value} = 1 - 0.9994 = 0.0006$$



Since the p-value is low, we reject H_0 . The data provide convincing evidence that more than 80% of Americans have a good intuition on experimental design.

Derivative of slides developed by Mine Çetinkaya-Rundel of OpenIntro.
Translated from LaTeX to Google Slides by Curry W. Hilton of OpenIntro.
The slides may be copied, edited, and/or shared via the
[CC BY-SA license](#)