Lecture 7:
Confidence Intervals for a Proportion
STAT 310, Spring 2021

Introduction

- A **point estimate** is our best guess of the value of a population parameter based on a random sample of data.
  - $\hat{p}$ is a point estimate p
  - $\bar{x}$ is a point estimate μ
- An **confidence interval** gives a range of plausible values for the population parameter.

Confidence Interval for p

A 95% confidence interval for the population proportion p is given by the formula

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\underset{\text{estimate}}{\text{point}} \pm 1.96\,SE$$

Confidence Interval for p

The confidence interval formula is valid if the following conditions are satisfied:

- The data were collected using simple random sampling. This is called the **independence** condition.
- $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$. This is called the **success-failure** condition.

These conditions ensure that the Central Limit Theorem holds, and so the sampling distribution for $\hat{p}$ is approximately normal.

Example

$\hat{p} = 0.56 \quad n = 1,202$

A recent Gallup poll estimated that 56% of Americans approved of the way Joe Biden is handling his job as president. The results were based on a random sample of 1,202 American adults.[1]

(a) Calculate and interpret a 95% confidence interval for the population proportion of American adults that approve of Joe Biden.

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow 0.56 \pm 1.96\sqrt{\frac{0.56(0.44)}{1202}}$$

$$\Rightarrow (0.532, 0.588)$$

We are 95% confident that the population proportion p of American adults that approve of Joe Biden is between 0.532 and 0.588

(b) Check the conditions for the interval.

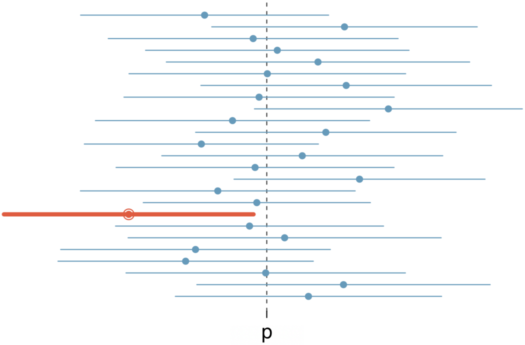
- $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$ ✓
- Data come from a random sample ✓

Yes, conditions are satisfied

1 https://news.gallup.com/poll/329948/biden-gets-high-marks-covid-response.aspx

What does 95% confidence mean?

Suppose we repeatedly took random samples of the same size from the population, and then constructed a 95% confidence interval using each sample. Then about 95% of those confidence intervals would contain the population proportion p.

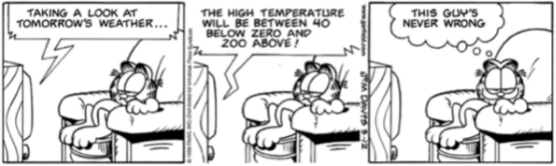


Width of an Interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

wider interval

Can you think of any drawbacks to using a wider interval?



If the interval is too wide it may not be very informative

Changing the Confidence Level

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

z^* is called the critical value, which depends on the confidence level (CL). Some common values are provided in the table below.

CL	z^*
90%	1.645
95%	1.96
99%	2.576

Example: Changing the Confidence Level $\hat{p} = 0.56, \; n = 1202$

A recent Gallup poll estimated that 56% of Americans approved of the way Joe Biden is handling his job as president. The results were based on a random sample of 1,202 American adults. Calculate a 99% confidence interval for the population proportion of American adults that approve of Joe Biden.

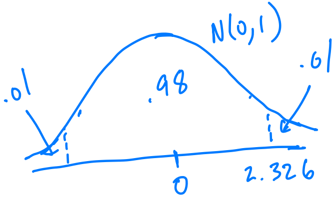
$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow 0.56 \pm 2.576 \sqrt{\frac{0.56(0.44)}{1202}}$$

$$\Rightarrow (0.523, 0.597)$$

Changing the Confidence Level

The value for the critical value z^* can be found manually using the R function `qnorm()`.

Example: Use R to find the critical value z^* that corresponds with a 98% confidence level.



$$z^* = \text{qnorm}(0.99)$$

$$= 2.326$$

Terminology

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

margin of error

- z^* is called the **critical value**, which depends on the confidence level
- $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is called the **standard error** (SE)
- $z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is called the **margin of error**

half width of the interval

Sample Size Determination

Determine the sample size needed so that the confidence interval will have a margin of error of $\pm E$

$$E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \implies E^2 = (z^*)^2 \frac{\hat{p}(1-\hat{p})}{n}$$

$$\implies \boxed{n = \frac{(z^*)^2 \hat{p}(1-\hat{p})}{E^2}}$$

- If no data has been collected then use $\hat{p} = 0.5$, which gives the largest possible sample size.
- When an estimate of the proportion is available, use it in place of 0.5.

Example: Sample Size Determination

A university newspaper is conducting a survey to determine what percentage of students support an increase in fees to pay for a new football stadium. How big of a sample is needed so that the margin of error is ± 0.04 using a 95% confidence level?

$E = 0.04$, $z^* = 1.96$, use $\hat{p} = 0.5$

$$n = \frac{(z^*)^2 \hat{p}(1-\hat{p})}{E^2} = \frac{(1.96)^2 (0.25)}{0.04^2}$$

$$= 600.25$$

A sample size of 601 is needed