Week 5 Confidence interval

2. Interpretation of confidence interval

Stat 140 - 04

Mount Holyoke College

Dr. Shan Shan

Slides posted at *http://sshanshans.github.io/stat140*

1. Yesterday: confidence interval

2. Main ideas
   1. What does 95% confidence really mean?
   2. How to compute the critical value
   3. How large a sample size do we need

In general, a confidence interval is of the form:

$$\text{Point Estimate} \pm \text{Critical value} \times \text{SE}$$

Critical value depends on

- Desired level of confidence (e.g. 90%, 95%, 99%)
- Appropriate sampling distribution (so far we have only learned normal distribution)

Suppose that in a random sample of $n = 200$ US residents who are 25 or older, 58 of them have a college bachelor's degree or higher. Construct a 95% confidence interval for the proportion of college graduates of US residents who are 25 or older.

- The point estimate is $\hat{p} = 0.29$
- We use $\hat{p} = 0.29$ to approximate p

$$\text{SE} = \sqrt{\frac{p(1-p)}{n}} = 0.014$$

- The 95% confidence interval is

$$\begin{aligned} \text{point estimate} \pm 2 \times \text{SE} &= 0.29 \pm 2 \times 0.014 \\ &= (0.262, 0.318) \end{aligned}$$

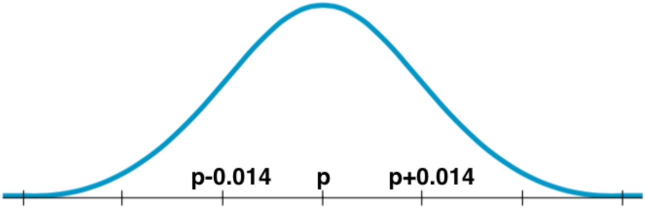
1. Yesterday: confidence interval

2. Main ideas
    1. What does 95% confidence really mean?
    2. How to compute the critical value
    3. How large a sample size do we need

1. Yesterday: confidence interval

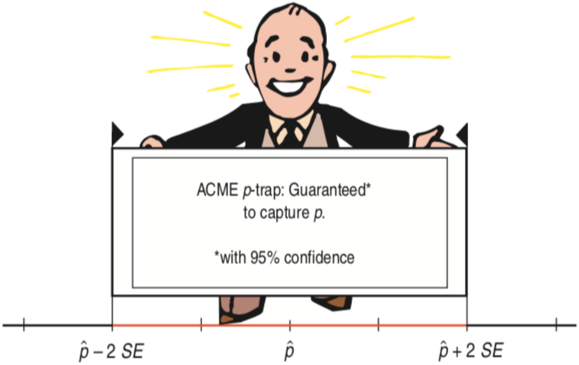
2. Main ideas
    1. What does 95% confidence really mean?
    2. How to compute the critical value
    3. How large a sample size do we need

Where does 95% of all samples fall within?



Where is our **observed** sample proportion in this picture?

Looking from $\hat{p}$'s perspective



If I'm $\hat{p}$, there's a 95% chance that p is no more than 2 SEs away from me. If I reach out 2 SEs away from me on both sides, I'm 95% sure that p will be within my grasp.

"95% confident that our interval contains the true proportion"

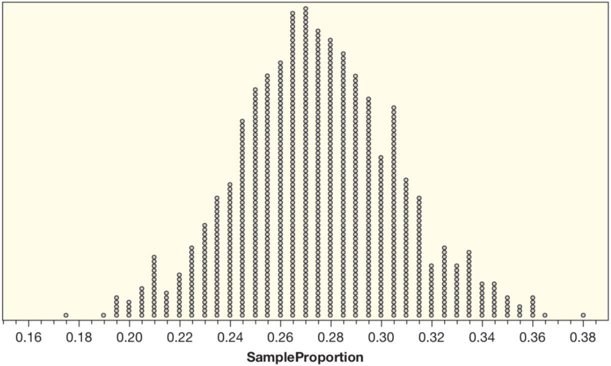
is equivalent to "95% of samples of this size will produce confidence intervals that capture the true proportion."

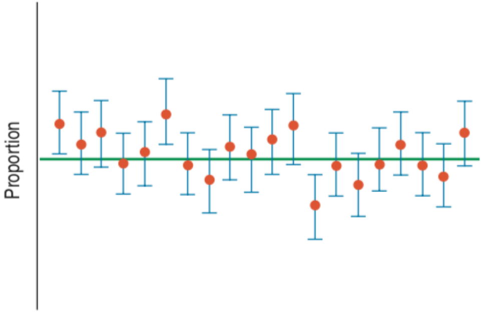
"95% confident that our interval contains the true proportion"

is equivalent to "95% of samples of this size will produce confidence intervals that capture the true proportion."

So what does this mean?

Imagine each dot of the picture below generates a confidence interval





The horizontal green line shows the true proportion of all US adults who are college graduates. Most of the 20 simulated samples produced confidence intervals that captured the true value, but a few missed.

We are 95% sure that our interval contains the true proportion.

- The point estimate is $\hat{p} = 0.29$
- The 95% confidence interval is $(0.262, 0.318)$

Can we say...

"proportion of college graduates of US residents who are 25 or older is 0.29."

- The point estimate is $\hat{p} = 0.29$
- The 95% confidence interval is $(0.262, 0.318)$

Can we say...

"proportion of college graduates of US residents who are 25 or older is 0.29."

No. In fact, we can be pretty sure that whatever the true proportion is, it's not exactly 0.29. So the statement is not true.

Also not true – "It is probably truly that the proportion of college graduates of US residents who are 25 or older is 0.29."

It would be nice to be able to make absolute statements about population values, but we just don't have enough information to do that.

- The point estimate is $\hat{p} = 0.29$
- The 95% confidence interval is $(0.262, 0.318)$

Can we say...

"We don't know exactly what is the true proportion of college graduates of US residents who are 25 or older, but **we know** that it's within the interval (0.262, 0.318)"

- The point estimate is $\hat{p} = 0.29$
- The 95% confidence interval is $(0.262, 0.318)$

Can we say...

"We don't know exactly what is the true proportion of college graduates of US residents who are 25 or older, but **we know** that it's within the interval (0.262, 0.318)"

This is getting closer, but it's still wrong. We just can't be certain. We can't know for sure that the true proportion is in this interval – or in any particular interval.

- The point estimate is $\hat{p} = 0.29$
- The 95% confidence interval is $(0.262, 0.318)$

Can we say...

"the true proportion of college graduates of US residents who are 25 or older, but the interval $(0.262, 0.318)$ **probably** contains the true proportion."

- The point estimate is $\hat{p} = 0.29$
- The 95% confidence interval is $(0.262, 0.318)$

Can we say...

"the true proportion of college graduates of US residents who are 25 or older, but the interval $(0.262, 0.318)$ **probably** contains the true proportion."

Now we're getting closer. We've admitted that we're unsure about two things. First, we need an interval, not just a value, to try to capture the true proportion. Second, we aren't even certain that the true proportion is in that interval, but we're "pretty sure" that it is.

"We are 95% confident that the true proportion of college graduates of US residents who are 25 or older lies within the interval $(0.262, 0.318)$"

Statements like these are the best we can do.

Misconception 1

A 95% confidence interval contains 95% of the data in the population.

Misconception 2

I am 95% sure that the proportion of a sample will fall within a 95% confidence interval for the population proportion.

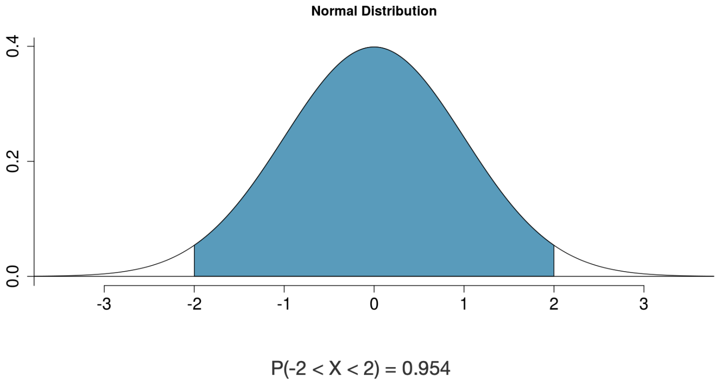
Misconception 3

The probability that the population parameter is in this particular 95% confidence interval is 0.95.

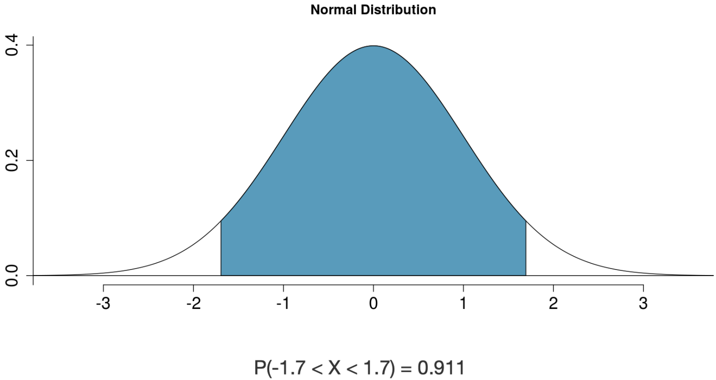
1. Yesterday: confidence interval

2. Main ideas
    1. What does 95% confidence really mean?
    2. How to compute the critical value
    3. How large a sample size do we need

For a 95% confidence interval, you'll find the precise critical value is $z^* = 1.96$. That is, 95% of a Normal model is found within 1.96 standard deviations of the mean. We've been using $z^* = 2$ because it's easy to remember.



For a 91.1% confidence interval, the critical value is 1.7, because, for a Normal model, 91.1% of the values are within 1.7 standard deviations from the mean.



1. Yesterday: confidence interval

2. Main ideas
    1. What does 95% confidence really mean?
    2. How to compute the critical value
    3. How large a sample size do we need

General Election: Trump vs. Biden (4-Way)

Top Battlegrounds: Trump vs. Biden

RCP Electoral Map | No Toss Up Map | RCP Senate Map | Senate NTU Map | RCP House Map | Latest Polls

Polling Data

Poll	Date	Sample	MoE	Biden (D)	Trump (R)	Spread
RCP Average	**10/16 - 10/26**	--	--	**50.6**	**43.5**	**Biden +7.1**
Emerson	10/25 - 10/26	1121 LV	2.8	50	45	Biden +5
IBD/TIPP	10/22 - 10/26	970 LV	3.2	50	46	Biden +4
Rasmussen Reports	10/22 - 10/26	1500 LV	2.5	49	47	Biden +2
JTN/RMG Research*	10/23 - 10/24	1842 LV	2.8	51	44	Biden +7
CNBC	10/21 - 10/24	800 RV	3.5	51	40	Biden +11
Economist/YouGov	10/18 - 10/20	1344 LV	3.2	52	43	Biden +9
Reuters/Ipsos	10/16 - 10/20	949 LV	3.6	51	42	Biden +9
Quinnipiac	10/16 - 10/19	1426 LV	2.6	51	41	Biden +10

All General Election: Trump vs. Biden Polling Data

Why is the margin of error smaller for the Emerson poll than the IBD/TIPP poll?

The question of how large a sample to take is an important step in planning any study.

$$\text{Margin of Error} = z^*\text{SE} = z^*\sqrt{\frac{p(1-p}{n}}$$

So if we know the desired ME, and confidence level (and hence z^*), and the SE, we can solve for n.

Suppose we want to determine a 95% confidence interval for voter support of Biden with a margin of error of no more than $\pm$ 3%. How large a sample do we need?

Let's use $\hat{p} = 0.5$ to estimate p.

$$\text{ME} = z^* \sqrt{\frac{p(1-p)}{n}}$$

$$0.03 = 1.96 \sqrt{\frac{(0.5)(0.5)}{n}}$$

To solve for n, we first multiply both sides of the equation by $\sqrt{n}$ and then divide by 0.03

$$0.03\sqrt{n} = 1.96\sqrt{(0.5)(0.5)}$$
$$\sqrt{n} = \frac{1.96\sqrt{(0.5)(0.5)}}{0.03} = 32.67$$

Therefore,

$$n = (32.67)^2 = 1067.$$