Inference for Single Proportions using the Normal Distribution

Stat 120

May 04 2023

Background

- Resampling inference methods like the bootstrap (CI) and randomization tests require the use of computers!
- We can achieve the same using statistical theory

Why are most resampling distributions bell-shaped?

CLT: when n is big enough, means and proportions behave like a normal distribution.

- Today we will compute SE using formulas derived from probability theory
- The inference methods in ch. 6+ are “classical” methods that could be done just with pen and paper.

The big question: Resampling vs. Classical methods

- *Resampling methods are intuitive and don't require lots of statistical theory/background.*
- *But in your research fields you will likely only see classical methods used*

- *In the "olden days", classical methods were the only thing taught in stats methods classes.*
- *More advanced methods usually do rely on classical theory due to their complexity.*

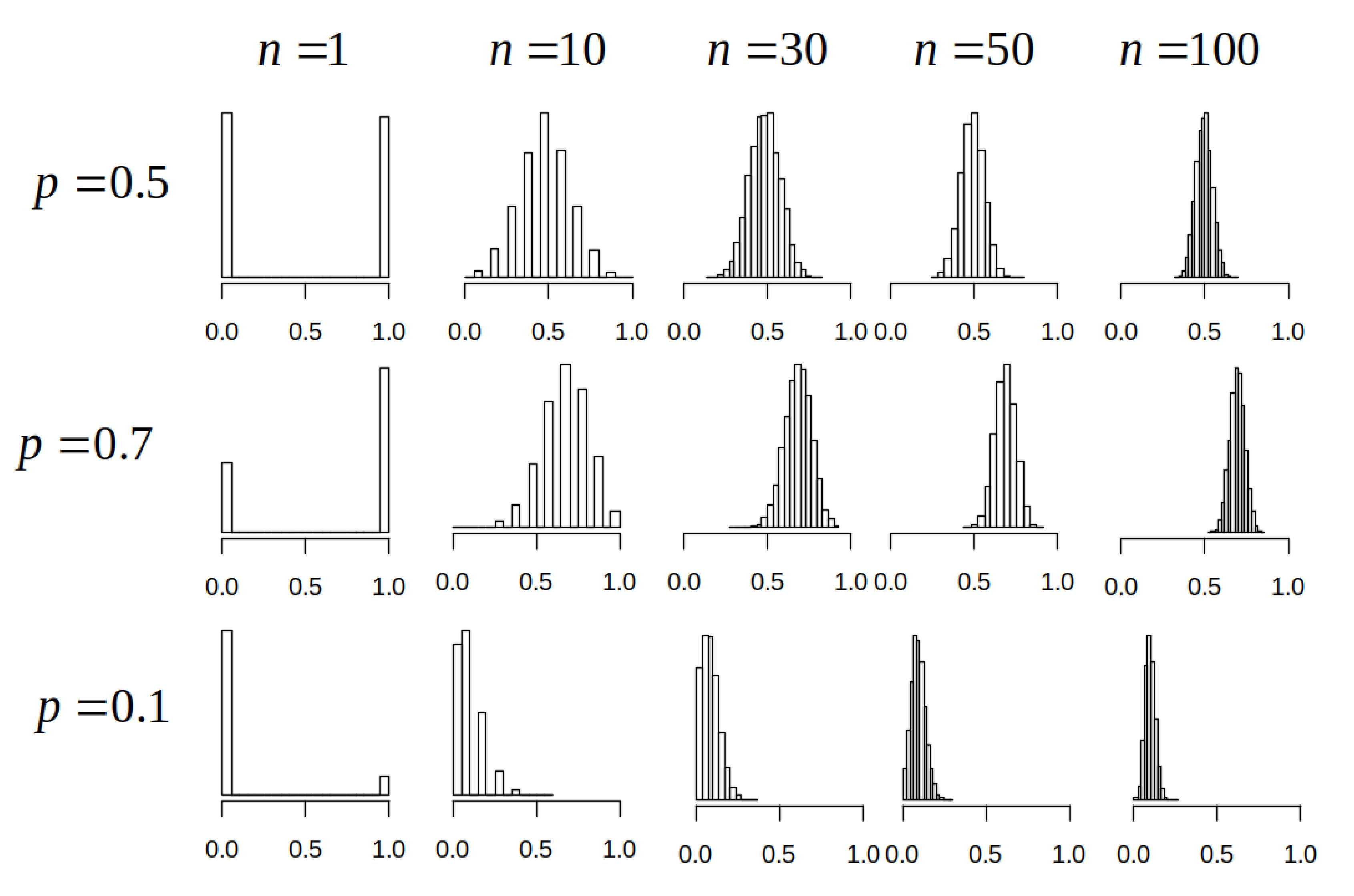
Quiz

The Central Limit Theorem applies to the distribution of the

1. *statistic*
2. *parameter*
3. *null value*
4. *data*
5. *standard error*

▶ Click for answer

Distribution of sample proportions



The SE for a Sample Proportion

The standard error for $\hat{p}$ is

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

The larger the sample size, the smaller the SE

Central Limit Theorem

For a sufficiently large sample size, the distribution of sample statistics for a mean or a proportion is normal

- *One sample proportion:* *The sampling distribution for a sample proportion is approximately normally distributed:*

$$\hat{p} \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Need n large enough so $np \geq 10$ and $n(1 - p) \geq 10$

Election polling

President Biden won 52.4% of the popular vote in Minnesota in the 2020 election.

- *If we had sampled 100 likely voters just prior to the election, what would be the SE for the sample proportion of voters for Biden?*

$$SE = \sqrt{\frac{0.524 \times 0.476}{100}} \approx 0.05$$

Margin of Error

For a single proportion, what is the margin of error?

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

1. $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
2. $z^* \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
3. $2 \times z^* \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

▶ Click for answer

Margin of Error and Sample Size

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

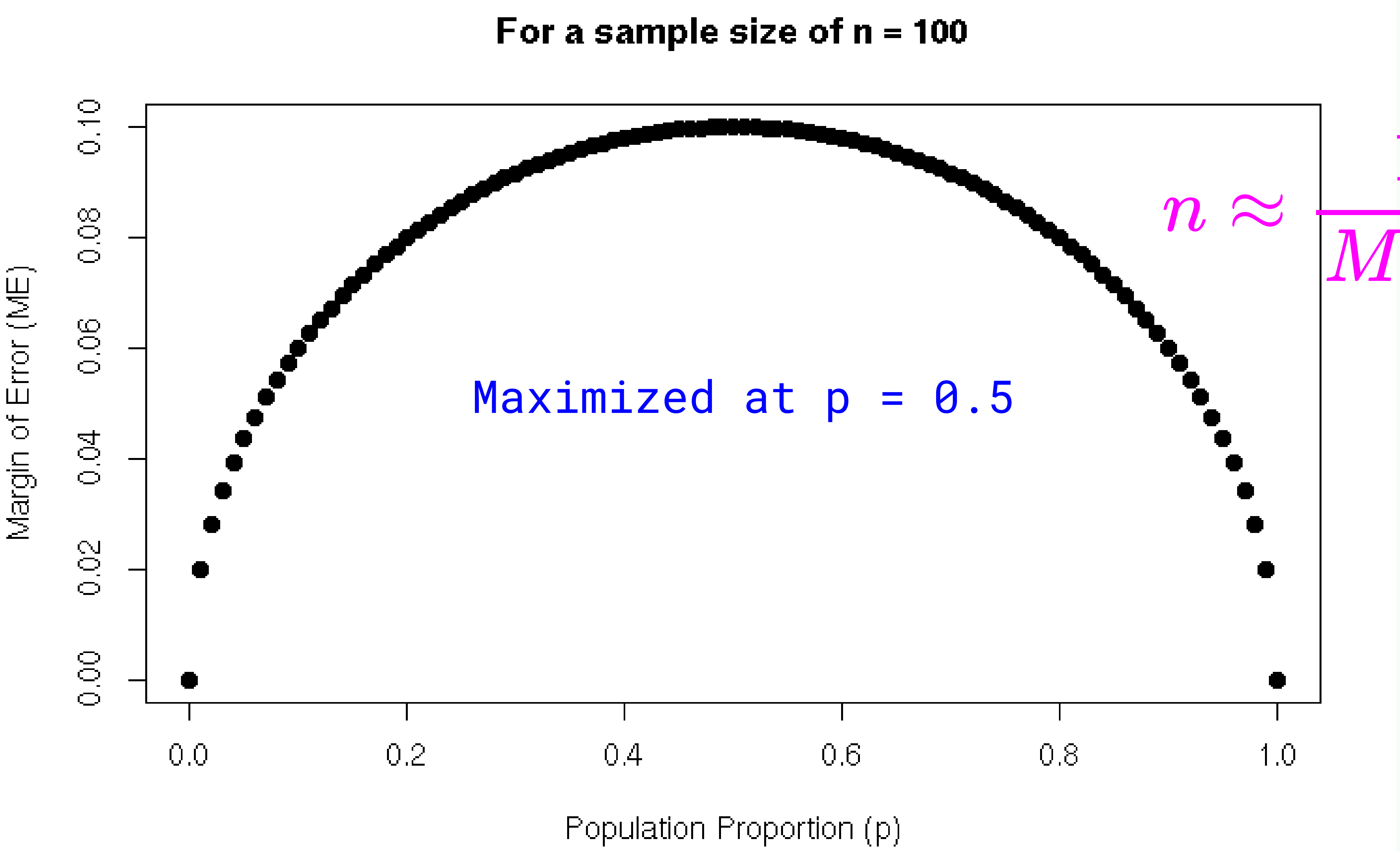
You can choose your sample size in advance, depending on your desired margin of error!

Given the formula for margin of error, solve for n.

- *Neither* p *nor* $\hat{p}$ *is known in advance. To be conservative, use* $p = 0.5$*. For a* 95% *confidence interval,* $z^* \approx 2$

$$n = \left(\frac{z^*}{ME}\right)^2 \hat{p}(1-\hat{p}) \qquad \Longleftrightarrow \qquad n \approx \frac{1}{ME^2}$$

Margin of Error and p



$$n \approx \frac{1}{ME^2}$$

Margin of Error and n: $n \approx \frac{1}{ME^2}$

Suppose we want to estimate a proportion with a margin of error of 0.03 with 95% confidence. How large a sample size do we need?

1. *About 100*
2. *About 500*
3. *About 1000*
4. *About 5000*

▶ Click for answer

Election polling continued..

What should n be to get a margin of error of 3%?

$$0.03 = 2 \times \text{ SE}$$

$$0.015 = SE = \sqrt{\frac{0.482 \times 0.518}{n}}$$

$$n = \frac{0.524 \times 0.476}{0.015^2} \approx 1109$$

Test for a Single Proportion: Standardized Test Stat and P-value

$$\begin{aligned} \mathrm{H}_0 : &\quad p = p_0 \\ \mathrm{H}_A : &\quad p \neq p_0 \end{aligned}$$

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

If $np_0 \geq 10$ and $n(1-p_0) \geq 10$, then the p-value can be computed as the area in the tail(s) of a standard normal beyond z.

Recap: Global Warming

Do a majority of Americans believe in global warming?

$$H_0 : p = 0.50$$
$$H_A : p > 0.50$$

$$p = \text{ proportion of all Americans who believe in global warming}$$

A survey on 2,251 randomly selected individuals conducted in October 2010 found that 1328 answered "Yes" to the question

"Is there solid evidence of global warming?"

Source: "Wide Partisan Divide Over Global Warming", Pew Research Center, 10/27/10 ↩

Is there solid evidence of global warming?

Sample proportion:

$$\hat{p} = \frac{1328}{2251} = 0.590$$

Standardized test stat:

$$z = \frac{0.590 - 0.50}{\sqrt{\frac{0.50(0.50)}{2251}}} = \frac{0.09}{0.0105} = 8.54$$

P-value:

```
1 - pnorm(8.54, 0, 1)
[1] 0
```

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

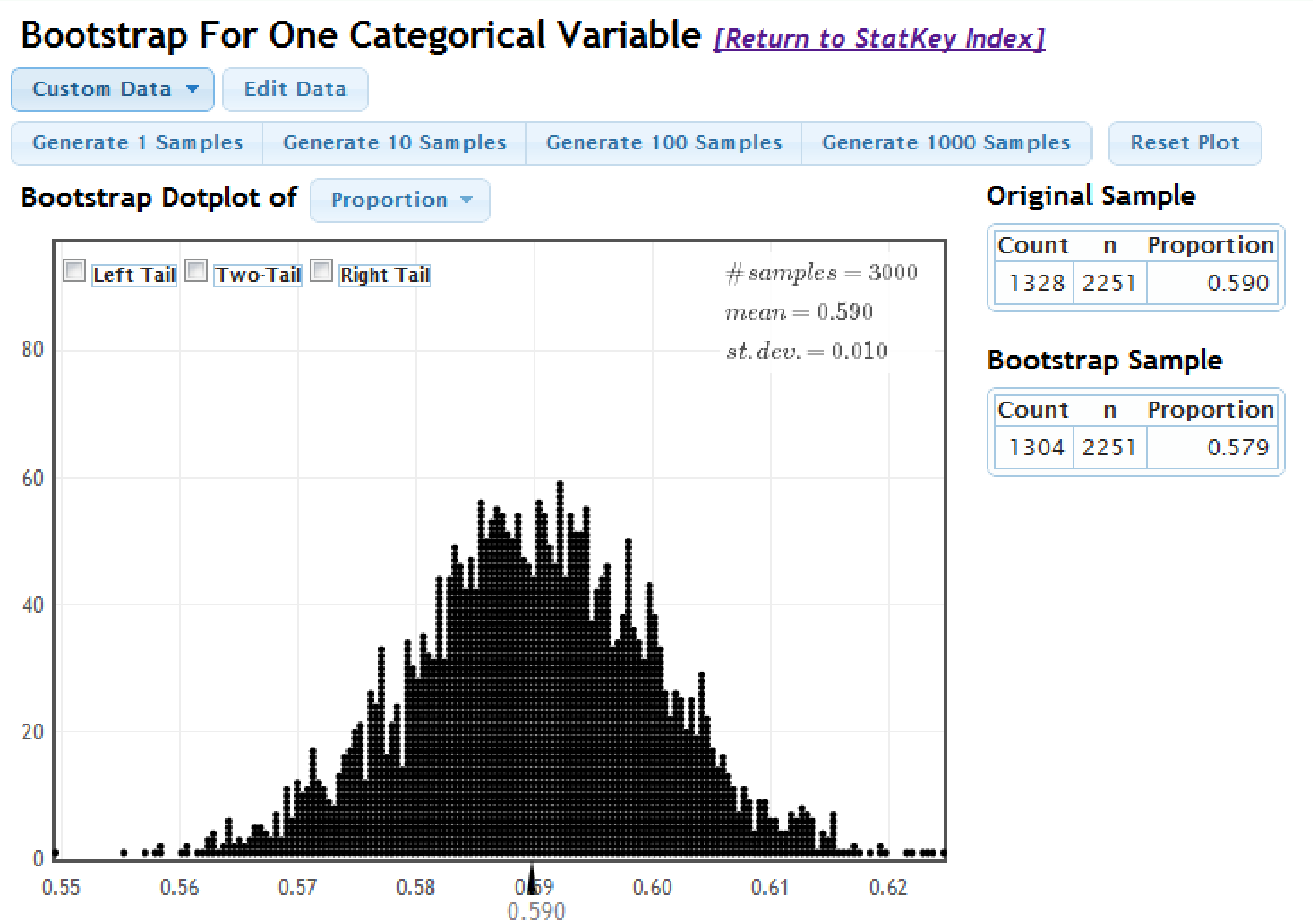
$$0.59 \pm 1.96 \sqrt{\frac{0.59 \times (1 - 0.59)}{2251}}$$

$$0.59 \pm 1.96 \times 0.0104 = (0.570, 0.610)$$

P-value: proportion above z=8.54 on a N(0,1) curve. Yes, there is strong evidence that the percentage of Americans that believe in global warming is greater than 50% (z=8.51, p<0.0001).

We are 95% confident that between 57% and 61% of Americans believe in global warming.

Statkey: Does this agree with the bootstrap CI?



We are *95%* sure that the true percentage of all Americans that believe there is solid evidence of global warming is between *57.0%* and *61.0%*.

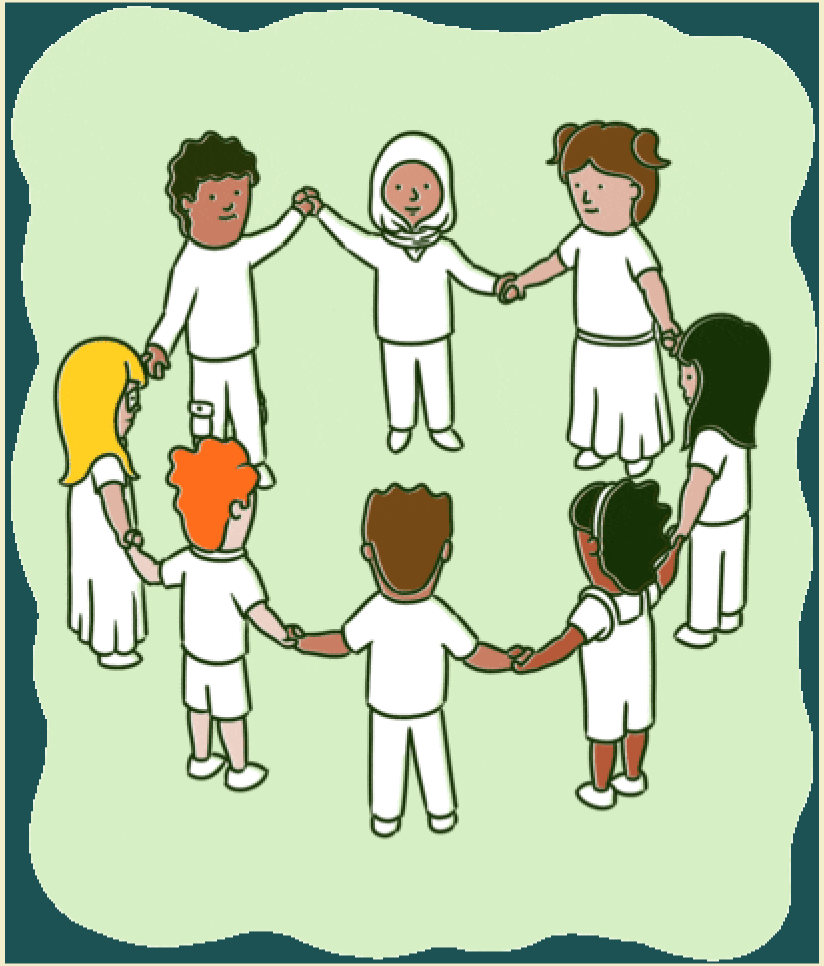
Summary

Standard error for a sample proportion: Central Limit Theorem for a proportion: If counts for each category are at least 10 (meaning $np \geq 10$ and $n(1-p) \geq 10$), then

- For a CI, use p-hat in place of p
- For a Hypothesis Test, use p_0 in place of p when calculating the standardized statistic

YOUR TURN 1

10:00



Let's go over to the class activity .Rmd file and complete the tasks for today.