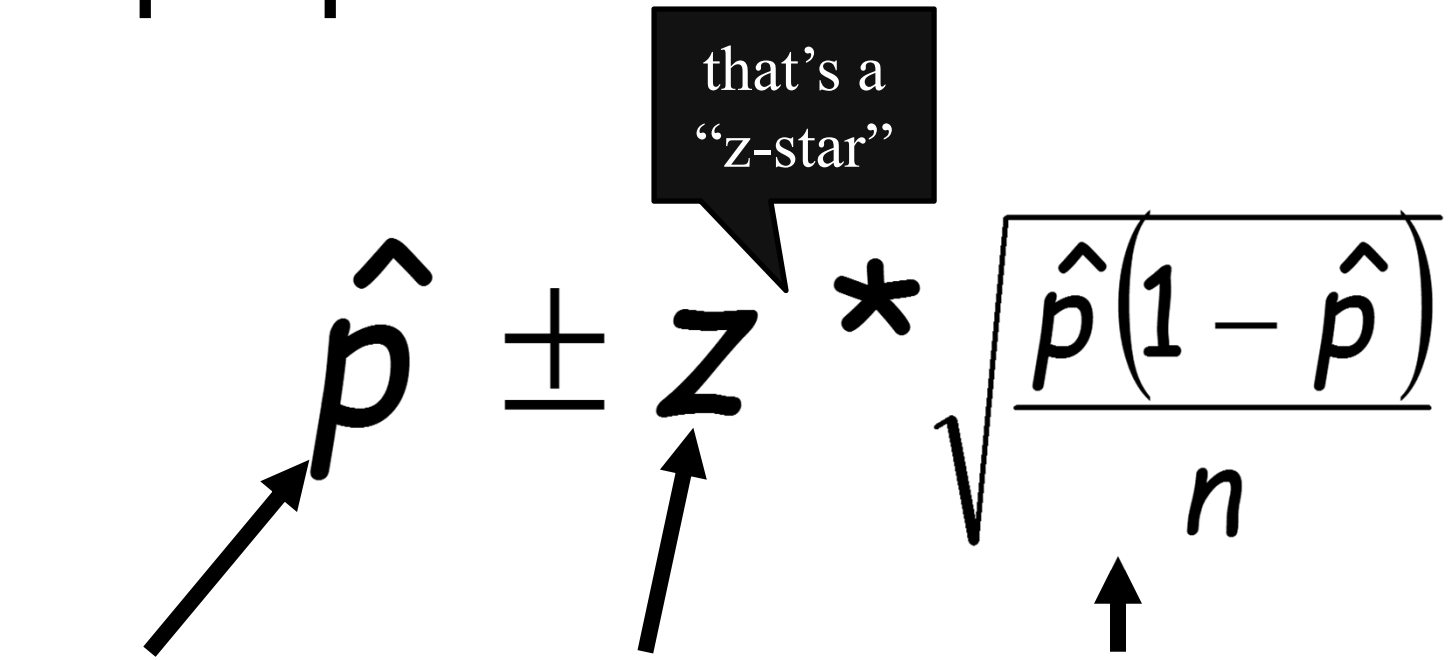
Confidence Intervals
with *proportions*

a.k.a., “1-proportion z-intervals”

AP Statistics
Chapter 19

1-proportion z-interval



Statistic ± Critical value × Standard deviation of the statistic

POINT ESTIMATE

STANDARD ERROR
(an estimate for std. dev)

CONDITIONS:

Check these EVERY TIME you construct a CI!!!

- Random (representative) sample?
  (so that we may generalize the results from the sample to the larger population)
- 10% rule?
  (so that sampling without replacement is similar to sampling with replacement – reasonable independence between selections)
- Successes/failures $\geq$ 10?
  (so that we may use the Normal model to approximate the distribution of p-hat)

Presidential Pete

A recent poll of 1020 randomly selected voters found 484 responding in favor of Pete. Construct a 95% confidence interval for the true proportion of all voters in Podunk who would vote for Presidential Pete.

Presidential Pete

A recent poll of 1020 randomly selected voters found 484 responding in favor of Pete. Construct a 95% confidence interval for the true proportion of all voters in Podunk who would vote for Presidential Pete.

p = true proportion of all voters that favor Pete

$$\hat{p} = \frac{484}{1020} = 0.4745 \qquad \hat{q} = 0.5255 \qquad n = 1020$$

1 sample *z*-interval for a proportion

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.4745 \pm 1.96\sqrt{\frac{(.4745)(.5255)}{1020}}$$

$$0.4745 \pm 0.0306$$

$$(0.4439,\ 0.5051)$$

Conditions:

- Given that we have a random sample of voters
- 1020 is reasonably less than 10% of all voters in Podunk.
- Normality: $n\hat{p} = 484 \geq 10$, $n\hat{q} = 536 \geq 10$

We are 95% confident that the true proportion of all voters in Podunk that favor Pete is between 44.4% and 50.5%

INTERPRETING A CONFIDENCE **INTERVAL**

(**ALWAYS** write this after finding a CI!!!)

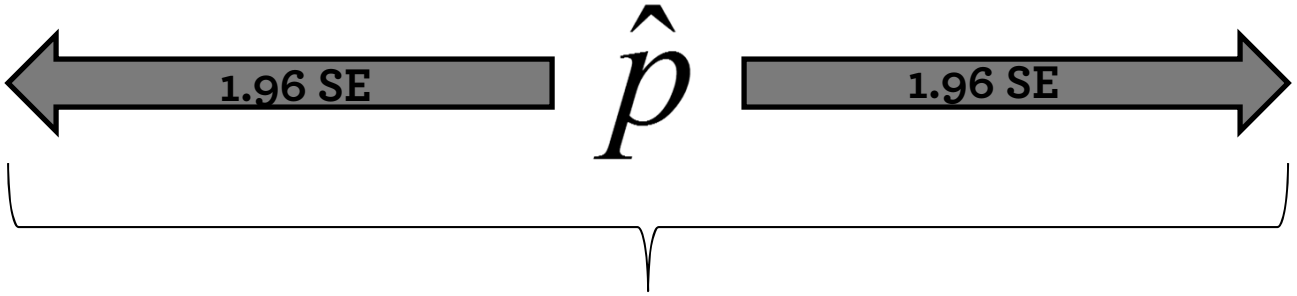
We are _____% confident that the true proportion of [“p” in context] is between _____ and _____.

“We are 95% confident that the true proportion of Podunkians who approve of Pete is between 0.4439 and 0.5051”

INTERPRETING A CONFIDENCE INTERVAL

(ALWAYS write this after finding a CI!!!)

"We are 95% confident that the true proportion of Podunkians who approve of Pete is between 0.4439 and 0.5051"



(We are 95% confident that "p" is in here somewhere)

INTERPRETING CONFIDENCE LEVEL

(only do this when asked)

If we collect a **LARGE** number of samples using this method, about _____% of the resulting confidence intervals will contain the true [proportion/mean in context]

(be sure to write in CONTEXT!!!)

*If we take a large number of samples using this method, about **95%** of the resulting confidence intervals will contain the true proportion of Podunkians who approve of Pete.*

confidence INTERVAL vs. confidence LEVEL

95% Confidence Interval: (0.4439, 0.5051)

Interpreting the Confidence <u>Interval</u>:

ALWAYS do this!!!

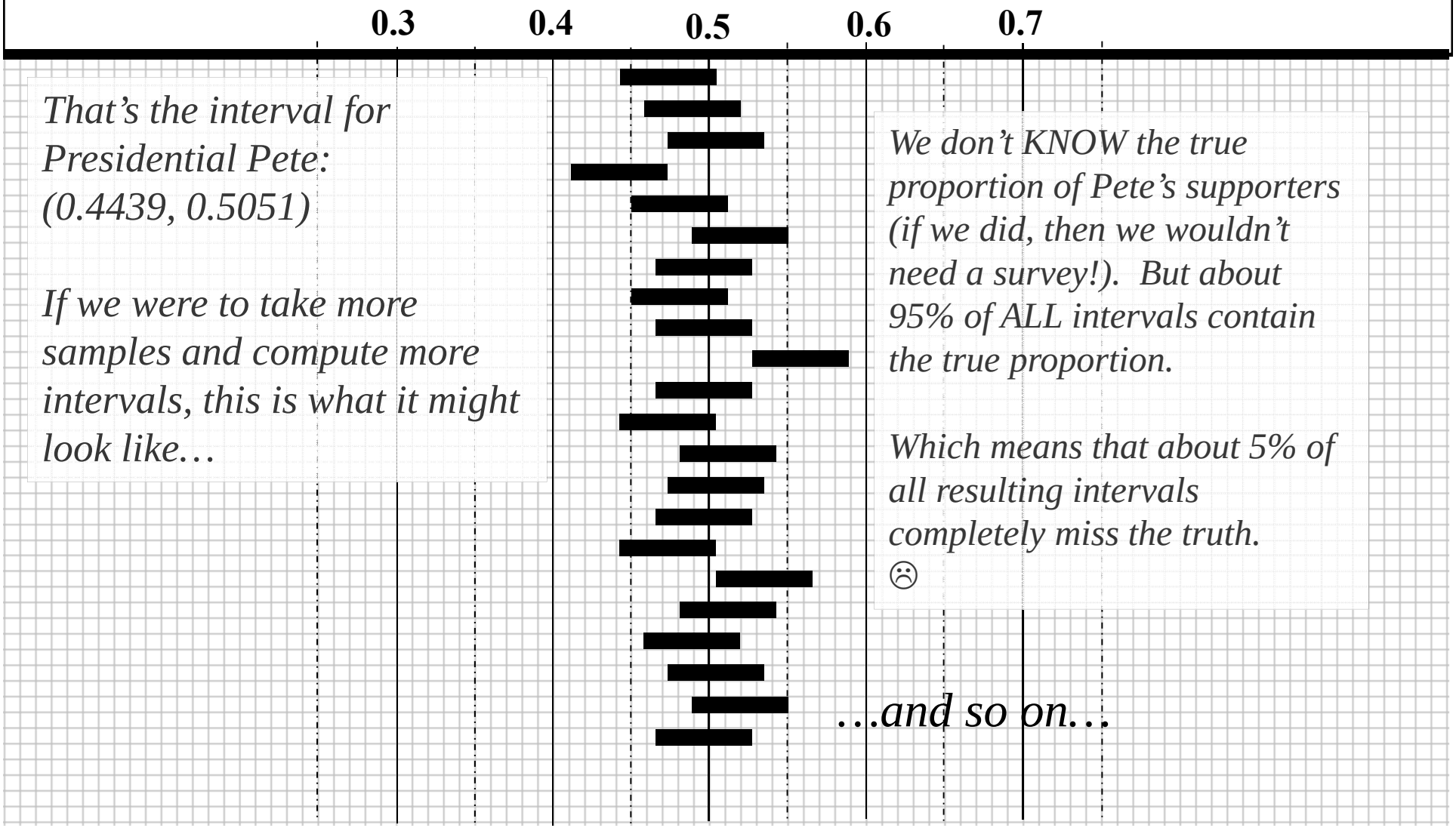
We are 95% confident that the true proportion of U.S. voters that favor Pete is between 44.4% and 50.5%

Interpreting the Confidence <u>Level</u>:

do this when asked.

Using this method, about 95% of all samples would produce an interval that captures the true proportion of all U.S. voters that are in favor of Pete.

The meaning of "95% Confidence"



That's the interval for Presidential Pete: (0.4439, 0.5051)

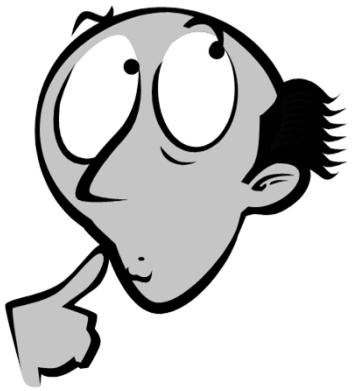
If we were to take more samples and compute more intervals, this is what it might look like…

We don't KNOW the true proportion of Pete's supporters (if we did, then we wouldn't need a survey!). But about 95% of ALL intervals contain the true proportion.

Which means that about 5% of all resulting intervals completely miss the truth. ☹

…and so on…

Rate your confidence
0 - 100



Guess my age within 10 years?

- **within 5 years?**
- **within 1 year?**

Catch a fly using a fly swatter?

- **using a pair of chopsticks?**
- **using a pair of toothpicks?**

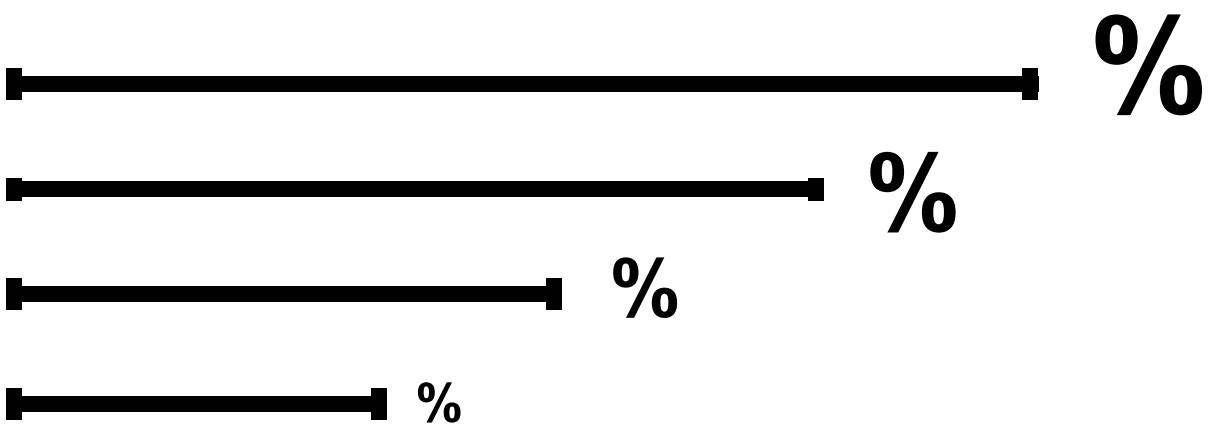
CONFIDENCE LEVEL, SAMPLE SIZE, MARGIN OF ERROR

As confidence level (z*) ↑, margin of error ↑

As sample size (n) ↑, margin of error ↓

$\times a$

$\times \frac{1}{\sqrt{a}}$



(Of course we need a good balance of confidence and precision...)

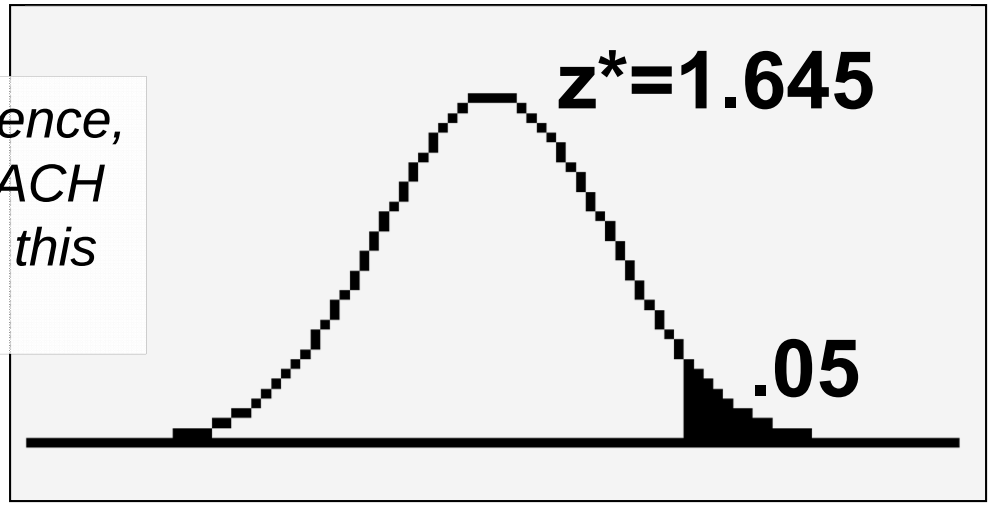
We are **100% confident** that the true proportion of Podunk voters who support Pete is **between 0% and 100%**!!!

Critical value (z*)

- **Divide the tail area in HALF**
- **Find the z-score for that percentile** *(take absolute value)*

Confidence level	tail area	z*
90%	.05	1.645
95%	.025	1.96
99%	.005	2.576

For instance, for 90% confidence, you cut off 5% (or 0.05) in EACH tail. The z-score to give you this is 1.645.



Hmmm… (questions to ponder)

1. What does the interval mean? Is Presidential Pete going to win??

   We don't know, because the interval contains values both above and below 50%

2. How would raising the confidence level to 98% change the interval?

   This makes the interval wider (the bigger the margin of error, the more confidence you have – think the bigger the fly swatter…)

3. Pete wants another poll taken but only has enough campaign money left to poll 500 voters. What would this change in sample size do to the interval?

   This also makes the interval wider – "n" is in the denominator of the formula – dividing by a smaller number makes the margin of error bigger/wider.

Once I get it, what does a CI mean?

1) We are 95% confident that the true proportion of voters in favor of Pete is between 44% and 50.5%.
2) There is a 0.95 probability that the true proportion of voters in favor of Pete is between 44% and 50.5%
3) If this procedure were repeated many times, approximately 95% of the sample proportions would be between 44% and 50.5%
4) The true proportion of all voters in favor of Presidential Pete is definitely between 44% and 50.5%.
5) If this procedure were repeated many times, approximately 95% of the resulting intervals would capture the true population proportion of voters in favor of Pete.
6) We are 95% confident that the sample proportion of voters in favor of Pete is contained in the interval (0.444, 0.505).

Only statements (1) and (5) are true. Anything that says "probability" or "percent of the time" for ONE specific interval… or that refers to "sample proportion" is missing the point…

"Mathematics is no more computation than typing is literature" – John Allen Paulos

Finding an appropriate sample size

Presidential Pete is on the home stretch of the election campaign, and wants to conduct a poll to estimate his level of support from potential voters. He wants to **ensure that the margin of error for the estimated proportion of voters that favor him is no more than ±0.02**. Find the necessary sample size, with 97% confidence.

$$\text{ME} = (\text{critical value}) \cdot (\text{standard error})$$

If you are reading this footnote... stop! focus on the problem!!!

The Human Cloning Problem

Another Gallop Poll is taken in order to measure the proportion of adults who approve of attempts to clone humans. **What sample size is necessary** to be within $\pm$ 0.04 of the true proportion of adults who approve of attempts to clone humans with a 95% confidence interval?

What p-hat (p) do you use when trying to find the sample size for a given margin of error?

$.1(.9) = .09$

$.2(.8) = .16$

$.3(.7) = .21$

$.4(.6) = .24$

$.5(.5) = .25$

By using **0.5 for p-hat**, we are using the worst-case scenario and using the largest SD in our calculations.

What sample size is necessary to be within ± 0.04 of the true proportion of adults who approve of attempts to clone humans with a 95% confidence interval?

$$m = z^* \left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

$$.04 = 1.96 \left(\sqrt{\frac{.5(.5)}{n}} \right)$$

Use p-hat = .5

$$\frac{.04}{1.96} = \sqrt{\frac{.5(.5)}{n}}$$

Divide by 1.96

$$\left(\frac{.04}{1.96} \right)^2 = \frac{.25}{n}$$

Square both sides

$$n = 600.25 \uparrow 601$$

Round up on sample size

High schoolers with cell phones

According to an article in the Chicago Tribune, about 75% of all high school students owned a cell phone in 2010. It is reasonable to believe that since that time, this proportion has increased. Suppose we wish to conduct a survey to estimate the current proportion of high school students that have a cell phone within 6% with 90% confidence. Find the necessary sample size.
(in this problem, we have additional information about the proportion of interest… so we should ***not*** *use 0.5 for p-hat)*

$$m = z^* \left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

$$.06 = 1.645 \left(\sqrt{\frac{.22(.78)}{n}} \right)$$

…

$$n = 140.9388 \uparrow 141$$