Math 247: Confidence Intervals for a Population Proportion *p* (Section 7.4)

Central idea: How do we use a Sample Proportion to estimate the value of a Population Proportion?

In order to investigate, we first have to understand what is meant by "Margin of Error"

Example: "Nutrition Facts" labels. HealthSmart makes a "100-calorie meal replacer" for people who are dieting. But the values given on the label can, by law, be off as much as 20 percent. Based on that, what could the actual calories be in a box of this stuff?

Nutrition Facts

7 servings per container
Servings size 1 packet (27g)

Amount Per Serving
Calories **100**

	% Daily Value
Total Fat 1g	2%
Saturated Fat 0g	0%
Trans Fat 0g	
Cholesterol 0mg	0%
Sodium 230mg	10%
Total Carbohydrate 6g	2%
Dietary Fiber <1g	4%
Total Sugars 3g	
Includes 0g Added Sugars	
Protein 15g	30%

"Margin of Error" = 20% of stated value
= 20 Cals

So 100 Cals is just an estimate

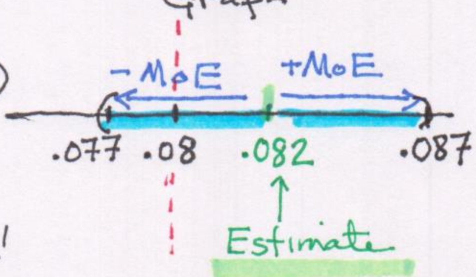
True value = Estimate ± M.o.E.
= 100 ± 20 Cals
= (100 − 20, 100 + 20)

True value is between 80 Cals and 120 Cals

Lab Tests and DUI's. Any time you have a lab test, the result has a Margin of Error. This includes lab tests for blood alcohol levels! Some states allow a defendant who is on trial for a DUI to use Margin of Error as part of the argument for their defense. Suppose a person has a lab test showing a blood alcohol level of .082 BAC (Blood Alcohol Concentration). If the test has a Margin of Error of .005, what would you think, if you were a juror on the case? Are we "beyond reasonable doubt" that the defendant truly was over the legal limit of .08 BAC?

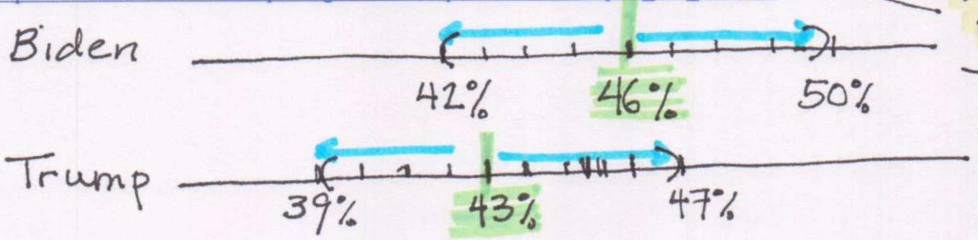
True value = Estimate ± M.o.E.
= .082 ± .005
= (.082 − .005, .082 + .005)
= (.077, .087)

It's possible this person's BAC was as low as .077, which is below the legal limit => don't convict!



Polls and Margin of Error: US Presidential Election, 2020. A poll was conducted in Arizona (an important swing state) in mid-March for a match-up between Joe Biden and Donald Trump for the upcoming presidential election. **The poll showed 46% of likely voters favoring Biden while 43% favored Trump.** That news was reported as "Biden leads Trump in Arizona", but if you read the fine print of the actual poll, they state the Margin of Error was ± 3.7% *with a 95% level of confidence*. Can we really say that Biden is ahead, based on this poll?
https://www.monmouth.edu/polling-institute/reports/monmouthpoll_AZ_031620/

What the heck is this???!



observe this overlap

Biden and Trump are in a "Statistical Tie" in Arizona, according to this poll!

* Note: We'll round MoE to 4% for convenience

A "**Confidence Interval**" in statistics is a range of values that "captures" the true population parameter we want to estimate. It's built out of an estimate (the "statistic", given by the sample data) combined with a "Margin of Error" (Note: This is more correctly called "Margin of Sampling Error")

The M.O.E. is derived from what we know about "sampling variability". The "Confidence Level" that we'll learn about next will also be based on the math behind sampling variability.

Confidence Interval for One Proportion

Margin of Error Format	Point Estimate ± Margin of Error $\hat{p} \pm \boxed{z^* \boxed{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}}$
Interval Format	$\left(\hat{p} - \boxed{z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}},\ \hat{p} + \boxed{z^* \boxed{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}}\right)$

The math behind a Confidence Interval and the "Confidence Level":

The Central Limit Theorem says that the Sampling Distribution of $\hat{p}$ will be approximately normal, provided the 3 conditions are satisfied

List the 3 conditions that a sample must meet for the Central Limit Theorem to apply:

1. Random Sample from population of interest Independent observations of subjects.
2. Large sample (Expected successes and failures more than 10)
3. Large population: Population size is at least 10 times the sample size.

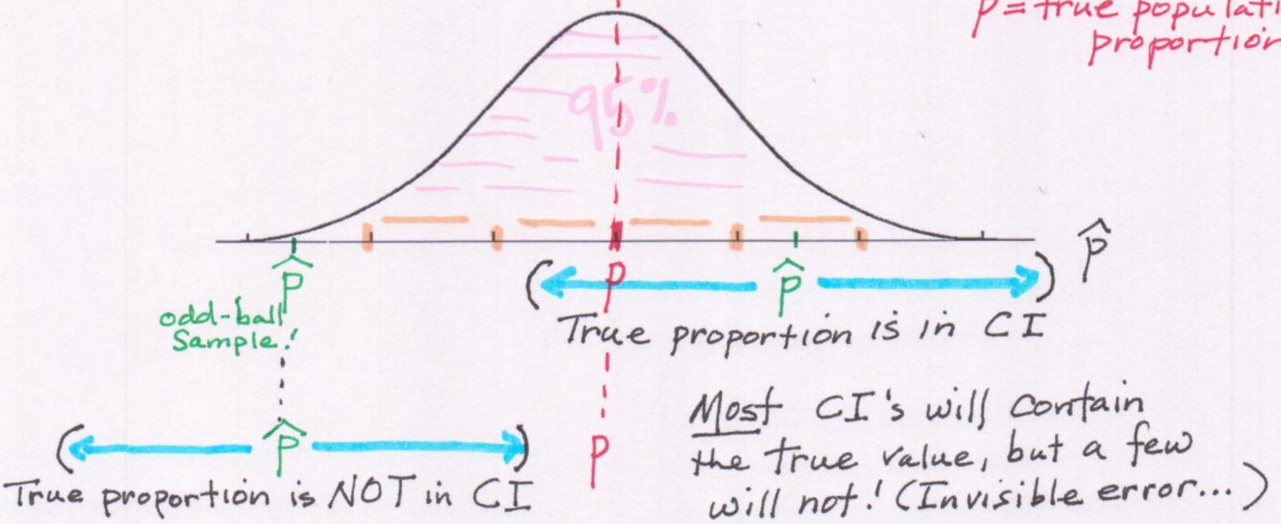
The standard deviation is called the "standard error" in this context. $SE = \sqrt{\frac{p(1-p)}{n}}$

Since we don't know the true population proportion, p, we have to estimate the SE value which is why the formula has $\boxed{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ in it. This is our best estimate for the SE.

Okay, from the Empirical Rule, we know that in a normal distribution, about 95% of the values in the distribution should be within 2 standard deviations of the mean. SE = standard deviation of a sampling distribution

The picture: About 95% of all the $\hat{p}$ values are between $p - 2 \cdot SE$ and $p + 2 \cdot SE$

p = true population proportion



We can reverse this and say that if we chose any particular value in the entire distribution, 95% of the time the mean would be within 2 standard deviations of that value.

About 95% of the time p will be located between $\hat{p} - 2 \cdot SE$ and $\hat{p} + 2 \cdot SE$ (or $\hat{p} \pm 2 \cdot SE$)

So, this interval contains ("covers" or "captures") the actual value of population proportion 95% of the time. We say that "we are 95% confident that p is in this interval".

But, oh, that 2 standard deviations value doesn't give us EXACTLY 95% (sigh), so we have to adjust it.

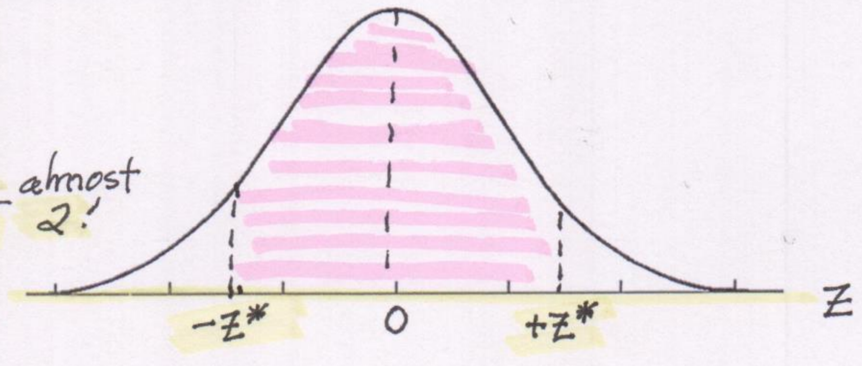
Remember that z-scores are measurements in terms of standard deviations (standard error in this context), so that "2" could be thought of as "z = 2"

We'll replace the "2" with a z-score, call it z*, that accurately cuts off the area under the normal curve that goes with the confidence level.

Here are the z-score values that go with common Confidence Intervals.

Summary of z^* values (fill in using the Normal Distribution Calculator on StatCrunch)

Confidence Level	z^*
80%	1.282
90%	1.645
95%	1.960 ← almost 2!
99%	2.576



Constructing a Confidence Interval

Example: Let's go back to that data from Iceland and use it to estimate the proportion of all COVID-positive people who are actually asymptomatic. They found that out of 180 people who tested positive, 90 of them (50 percent!) did not have any symptoms.

By hand, construct and interpret a 90% confidence interval for the proportion of all U.S. adults who are worried about the environment. Write the confidence interval in both formats.

Parking Lot

$n = 180$ positive people (Population of interest)

$x = 90$ no symptoms

$\hat{p} = \frac{x}{n} = \frac{90}{180} = .50$

90% confidence level $\Rightarrow z^* = 1.645$

Construct:

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

$= .50 \pm 1.645 \sqrt{\frac{.5(1-.5)}{180}}$

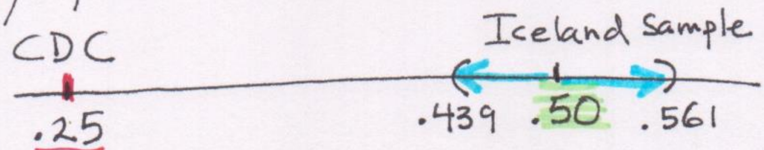
MoE format: $= .50 \pm .061$

$= (.50 - .061, .50 + .061)$

Interval format: $= (.439, .561)$

Interpret:

We are 90% confident that between 43.9% and 56.1% of COVID-positive people are asymptomatic, based on the data from Iceland.



Note: 90% isn't very high confidence. Check 99% confidence ~~don~~ using StatCrunch.

Confirm the CI using StatCrunch. Go through the same steps as for a hypothesis test, then select "Confidence Interval for p" instead of "Hypothesis Test for p", fill in the Confidence "Level" and leave the Method as "Standard-Wald".

Fill in the results below:

90% confidence interval results:

CI

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	90	180	0.5	0.037…	.43869992 (.439	.56130008 .561)

99% CI:
(.404, .596)