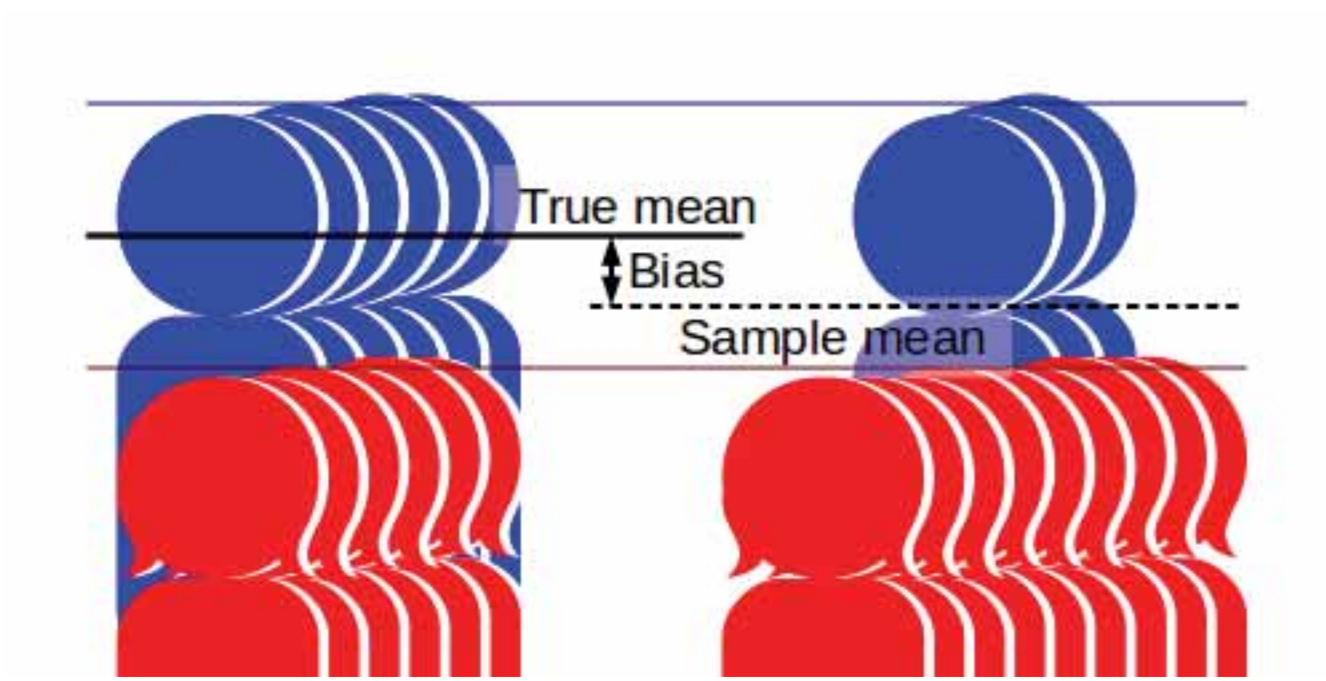
Raven, convinced that more of her classmates need to understand that Patrick Mahomes is the GOAT, takes a random survey of 50 SI students to see how many believe that Mahomes is, in fact, the GOAT. Her results are that 31 believe he is.

Given that this is one random sample, what can we say about the actual proportion of SI students who believe that Mahomes is the GOAT based just on Raven's random sample?

Biased vs. Unbiased Statistics



Statistic

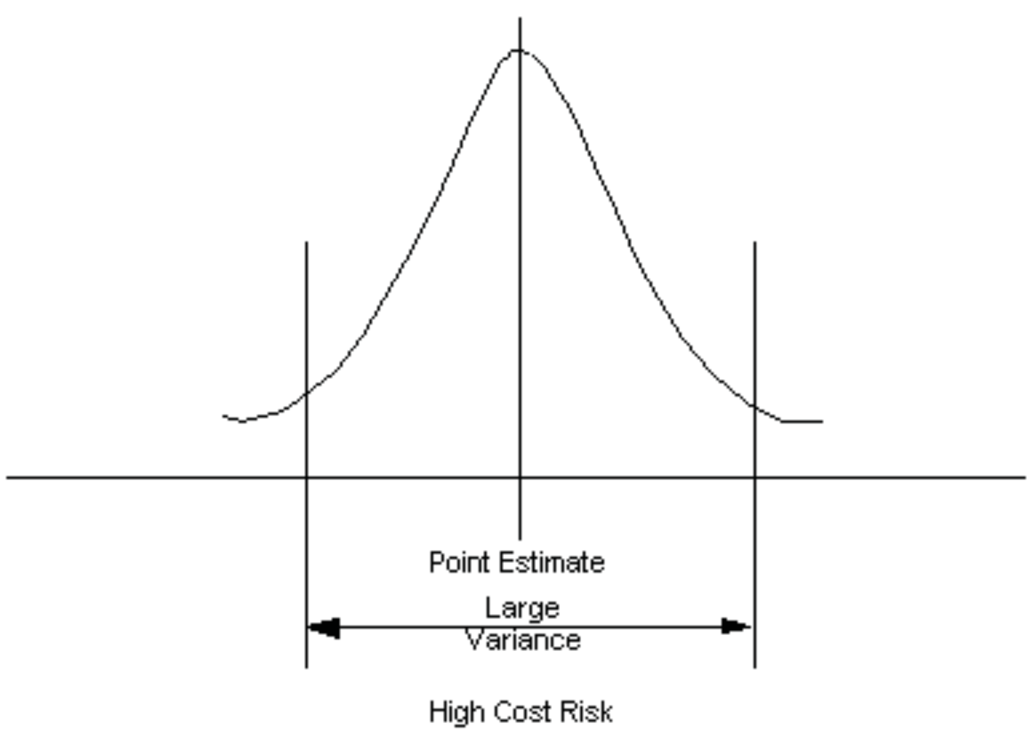
$\equiv$

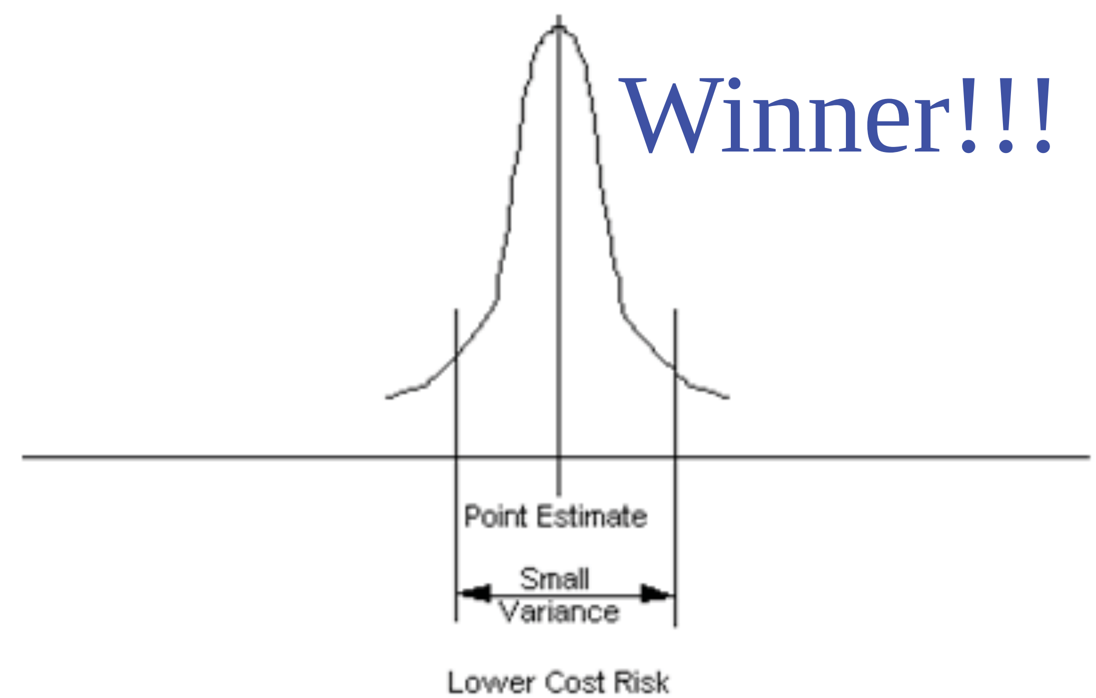
Point Estimate

$\hat{p}$ or $\bar{x}$

Sample proportion

Sample mean





Winner!!!

Point Estimate = one value to estimate the parameter based on sample data (we've called these **statistics** all year).

Confidence Intervals = range of values to estimate the parameter

We use our point estimate (our sample mean or sample proportion) to construct our confidence interval

Developing a CI involves using z scores so let's try a little algebra on this

$$z = \frac{\bar{x} - \mu}{\sigma}$$

$$z\sigma = \bar{x} - \mu$$

$$\mu = \bar{x} - z\sigma$$

Since z can be positive or negative and the true mean μ can be greater than or less than our sample mean, we can write this:

$$\mu = \bar{x} \pm z\sigma$$

So our confidence interval for the point estimate would be between these two values:

$$\bar{x} \pm z\sigma$$

Since we'll be looking at proportions in this unit, we'll use this interval:

$$p = \hat{p} \pm z\sigma$$

p = the true proportion of a population

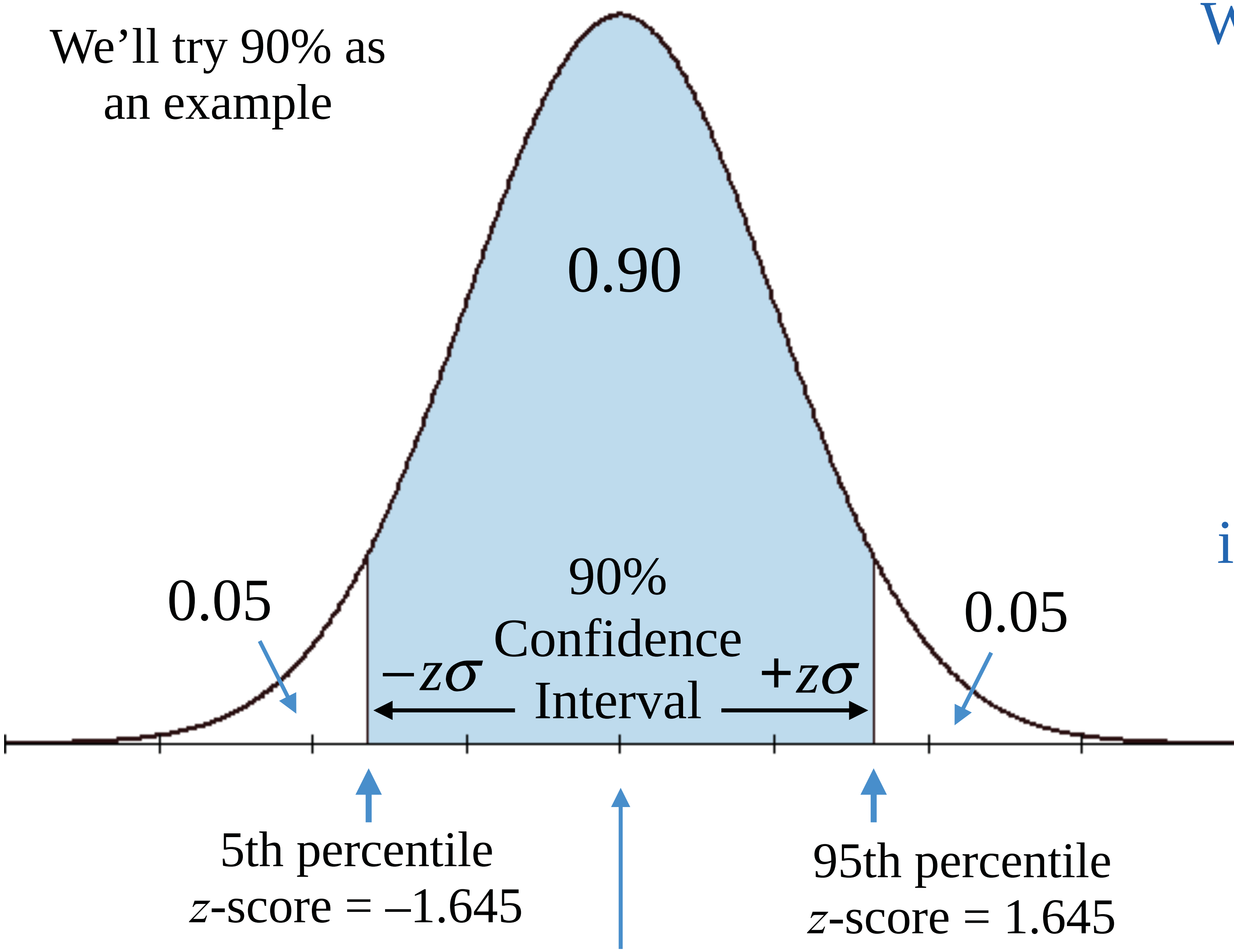
$\hat{p}$ = the sample proportion

Let's get a visual of this

Point Estimate = one value to estimate the parameter based on sample data (we've called these **statistics** all year).

Confidence Intervals = range of values to estimate the parameter

We'll try 90% as an example



5th percentile
z-score = –1.645

95th percentile
z-score = 1.645

Your point estimate

$\bar{x}$ or $\hat{p}$

We are 90% confident that

the true value of p
(population proportion)

or

the true value of μ
(population mean)

is within the given interval

Remember, this all started with z-scores

$$\mu = \bar{x} \pm z\sigma$$

$$p = \hat{p} \pm z\sigma$$

Confidence Intervals

General CI Formula

$$\text{Statistic} \pm (\text{Critical Value})(\text{Standard Deviation})$$

Let's start with the sample proportion confidence interval: $\hat{p} \pm z\sigma$

1 Sample Proportion CI Formula

Use Table or Calculator to get the z critical value

Notice the s.d. of the sample proportion?

$$\hat{p} \pm \underbrace{z\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}}$$

This is called the Margin of Error

Here are three *z* values to remember…

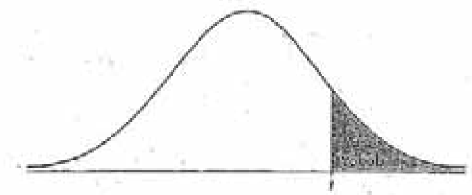
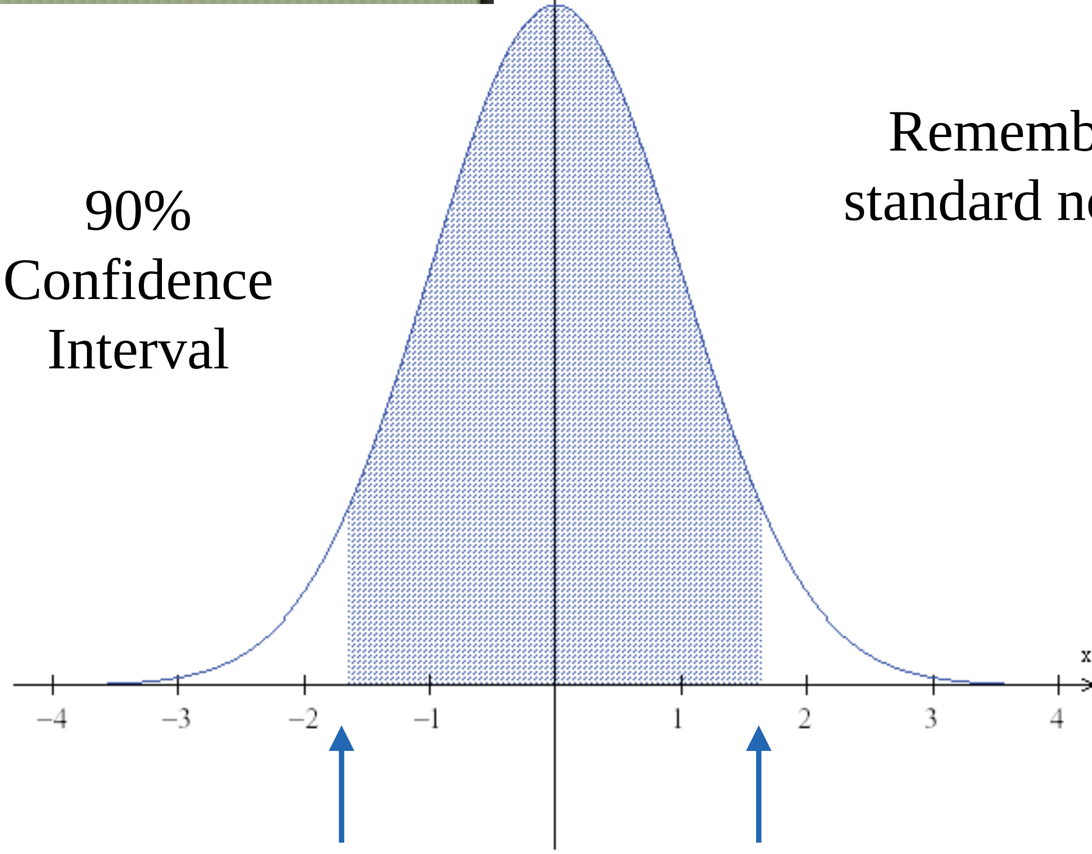
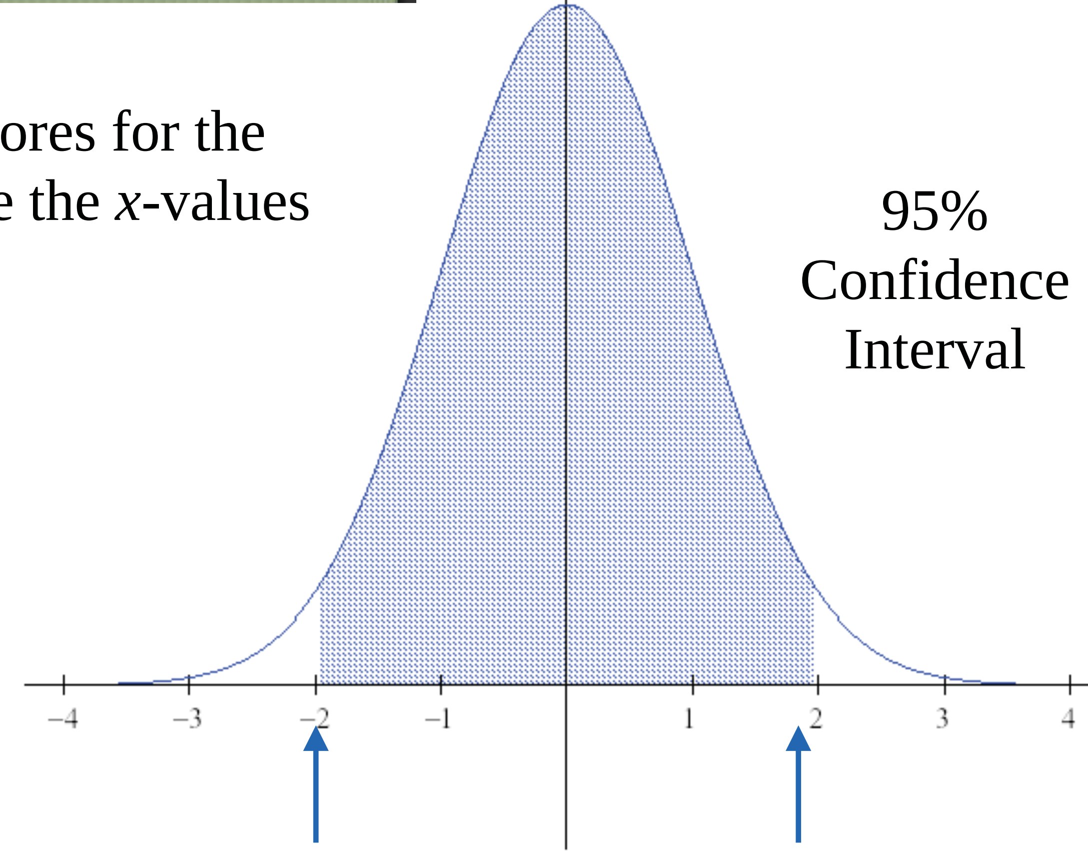


TABLE B: t-DISTRIBUTION CRITICAL VALUES

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											



Remember that the z scores for the standard normal curve are the *x*-values



90% Confidence Interval

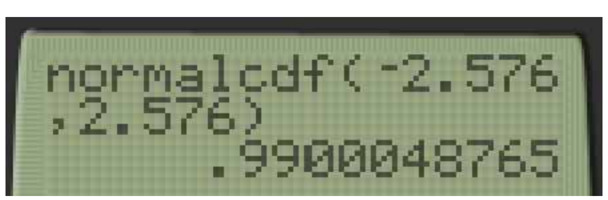
95% Confidence Interval

−1.645 1.645

−1.96 1.96

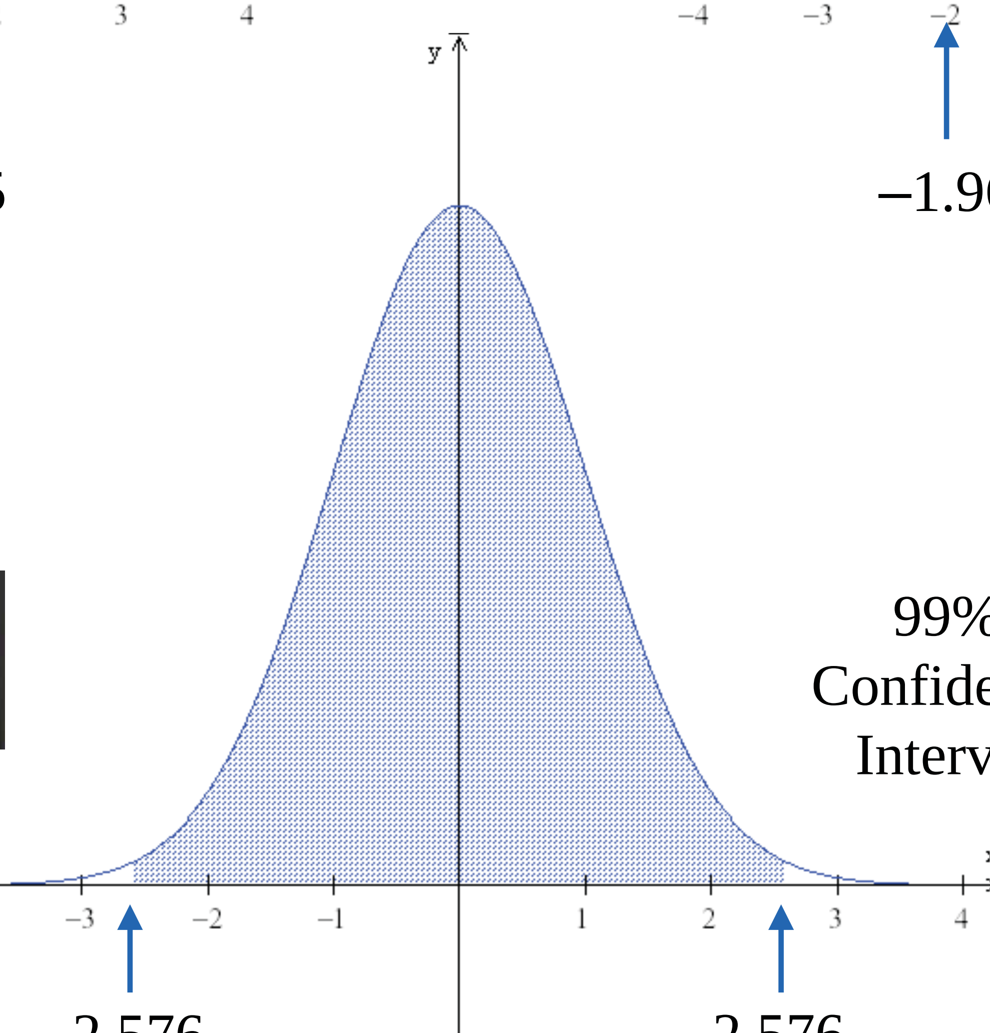
$z = x = \pm 1.645$

$z = x = \pm 1.96$

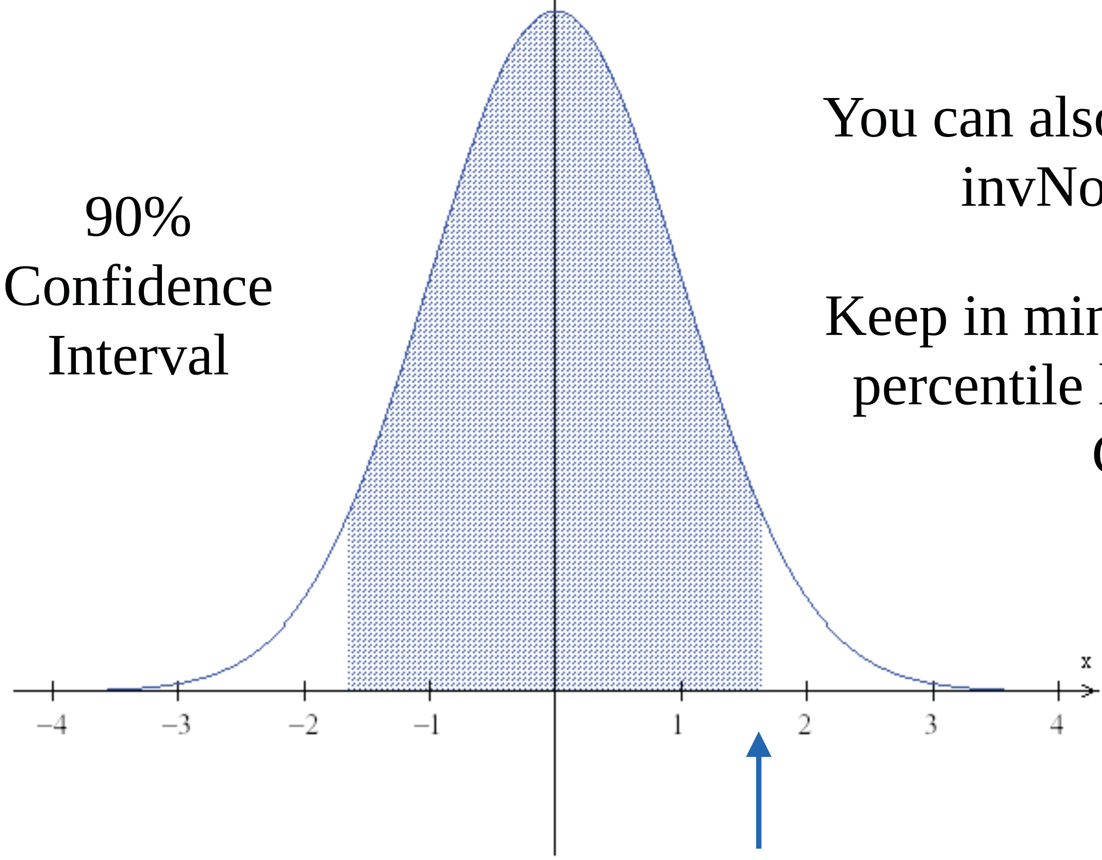


99% Confidence Interval

$z = x = \pm 2.576$



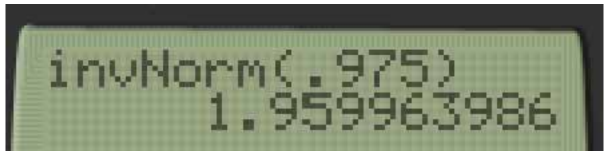
−2.576 2.576

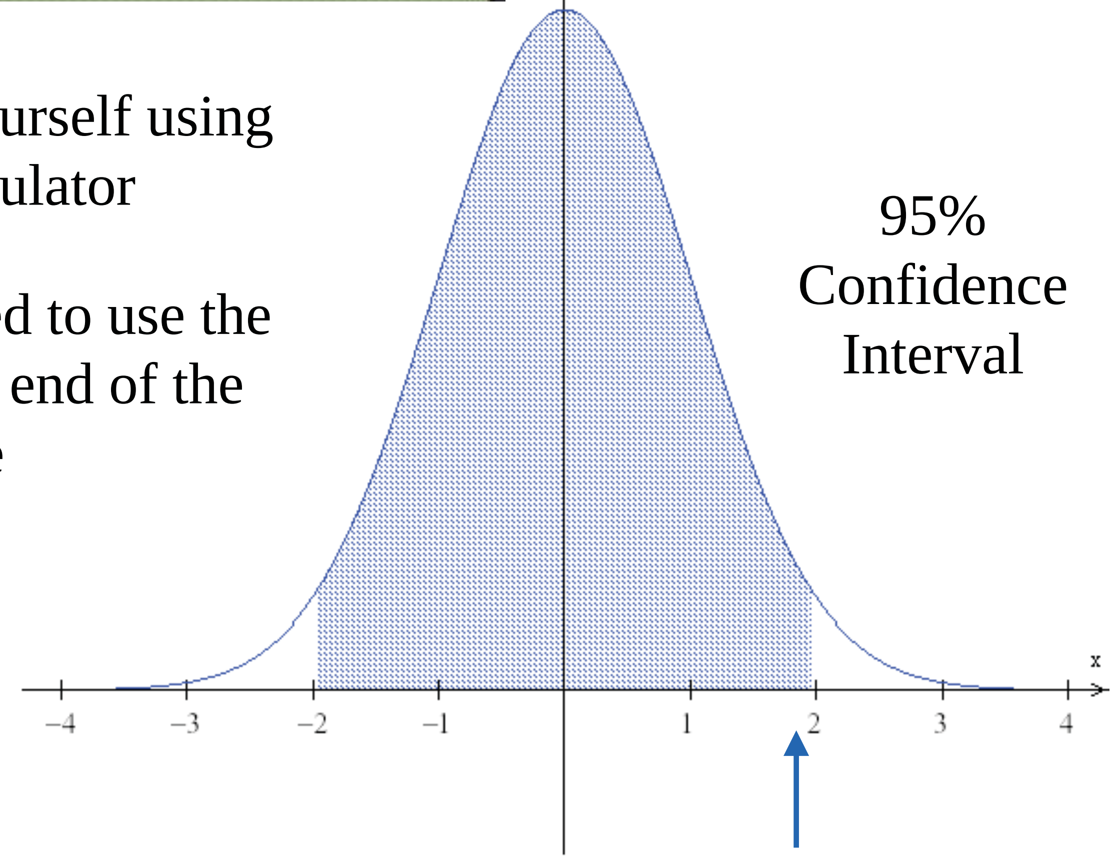


90% Confidence Interval

You can also find these yourself using invNorm on the calculator

Keep in mind that you need to use the percentile location of the end of the CI, not the size





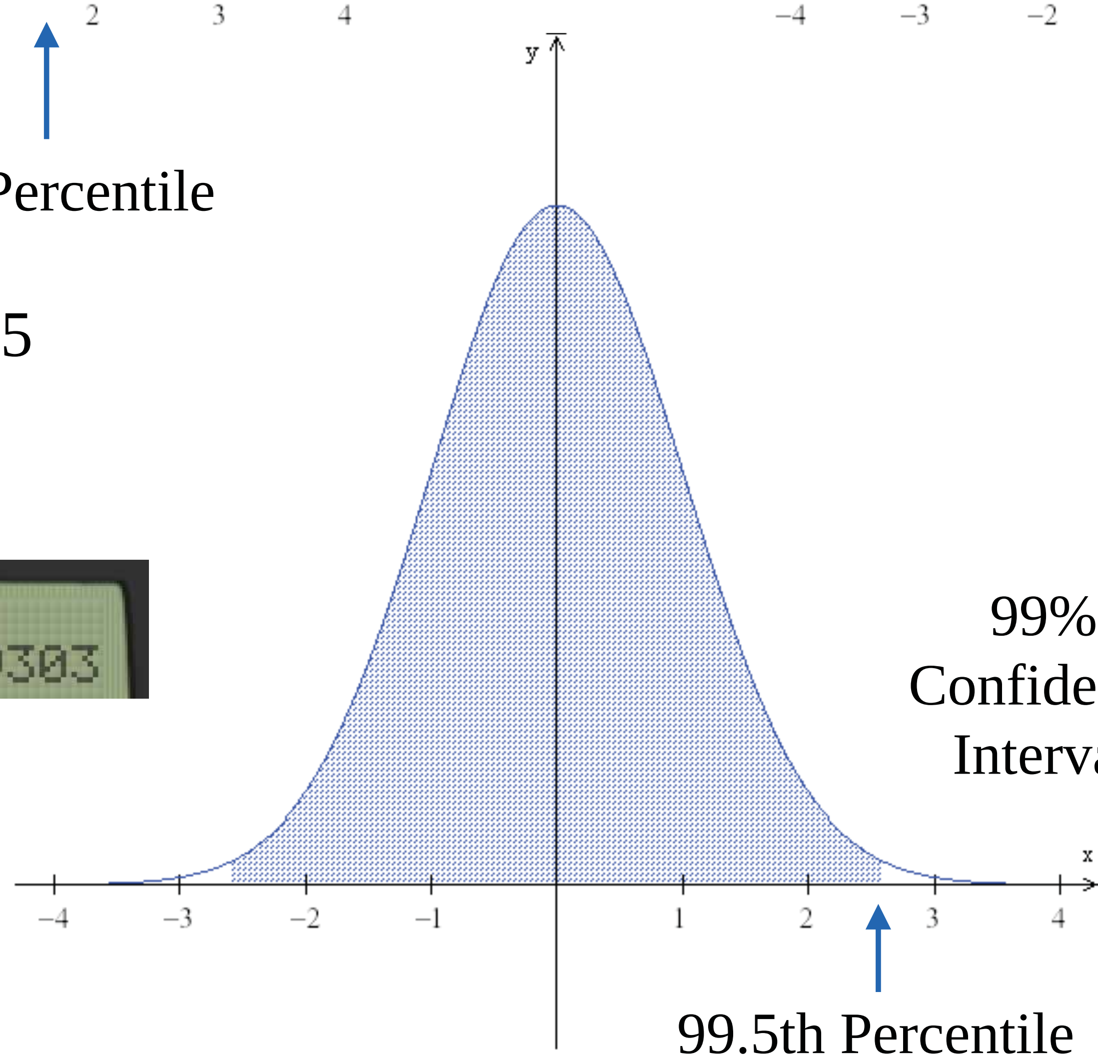
95% Confidence Interval

95th Percentile

97.5th Percentile

$z = x = \pm 1.645$

$z = x = \pm 1.96$



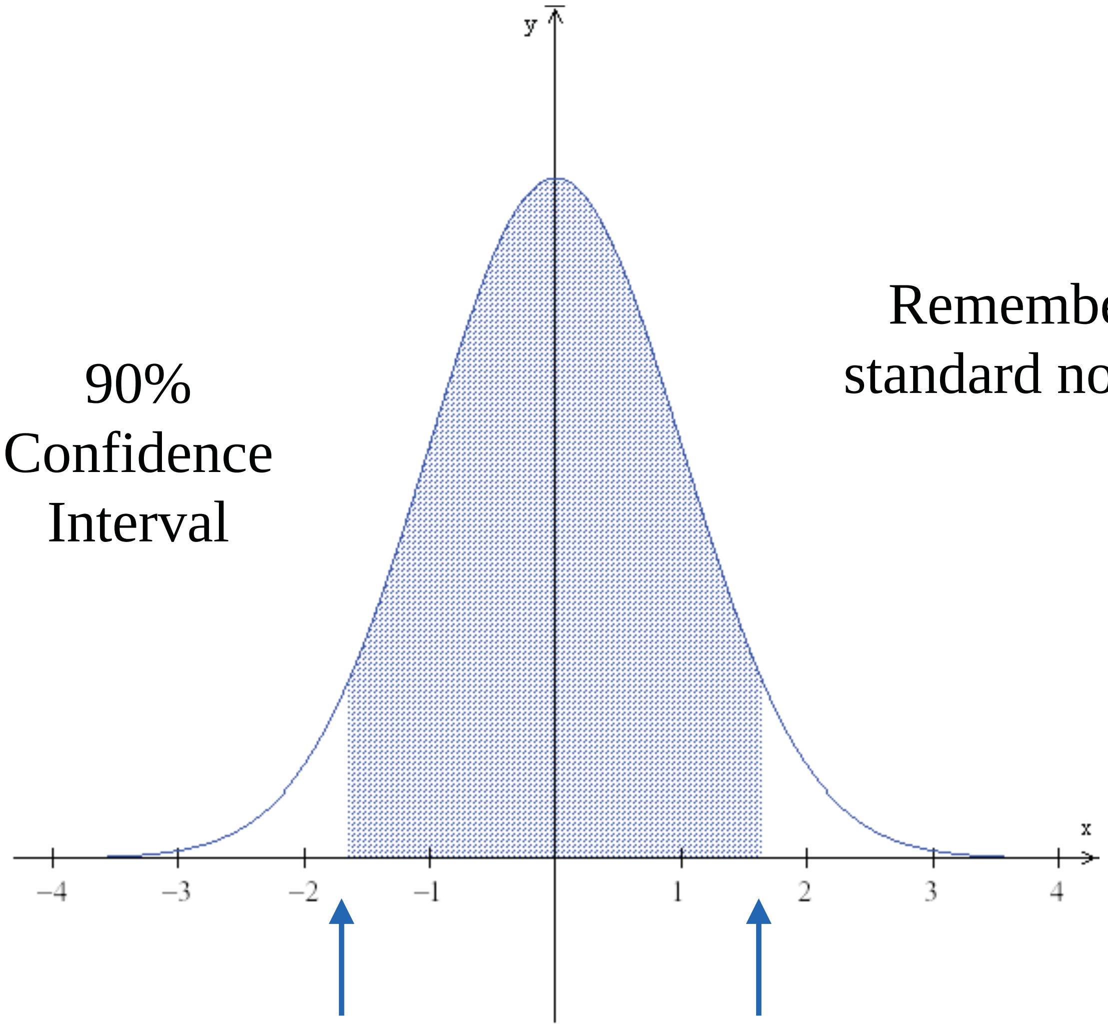
99% Confidence Interval

$z = x = \pm 2.576$

99.5th Percentile

Remember that the z scores for the standard normal curve are the *x*-values

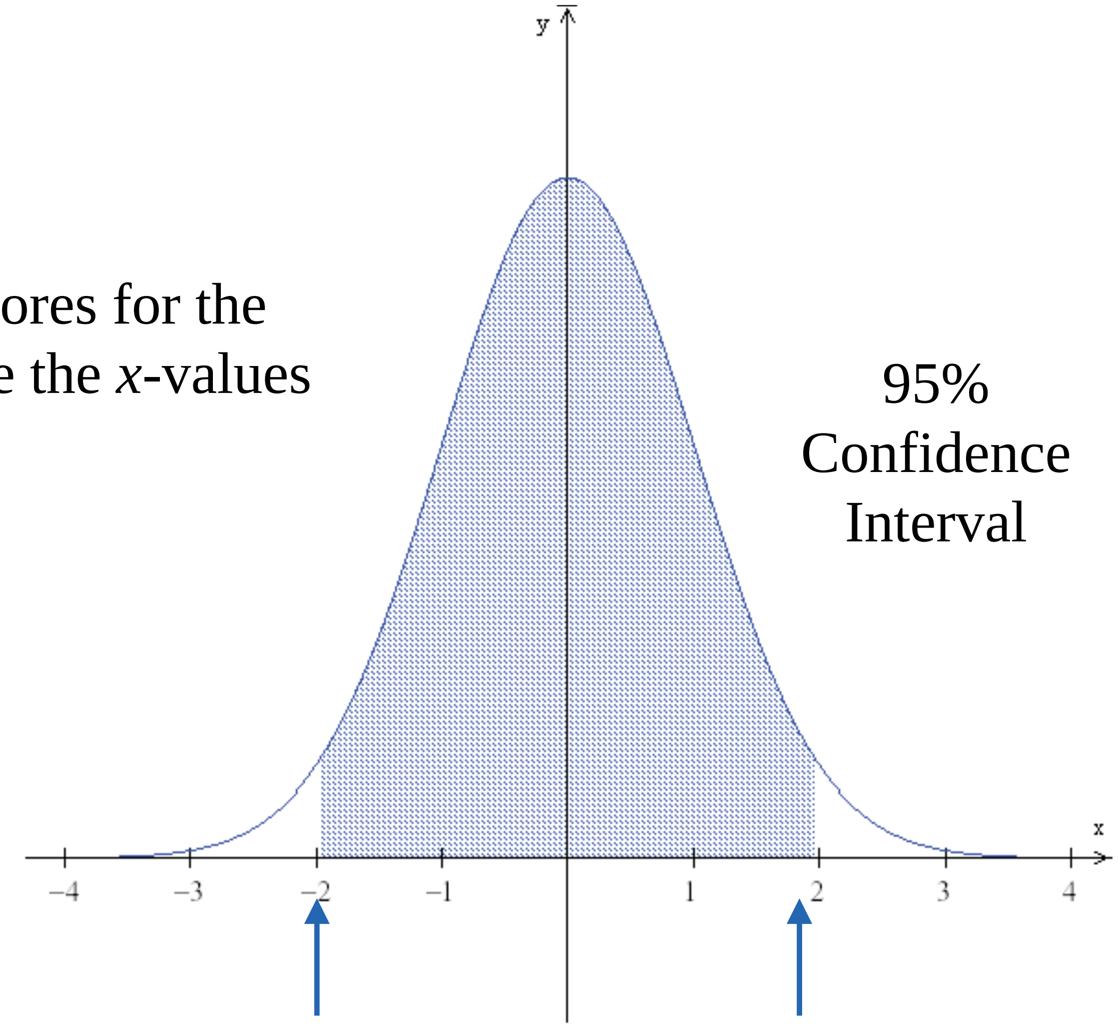
90%
Confidence
Interval



-1.645 $\qquad$ 1.645

$$\hat{p} \pm 1.645\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

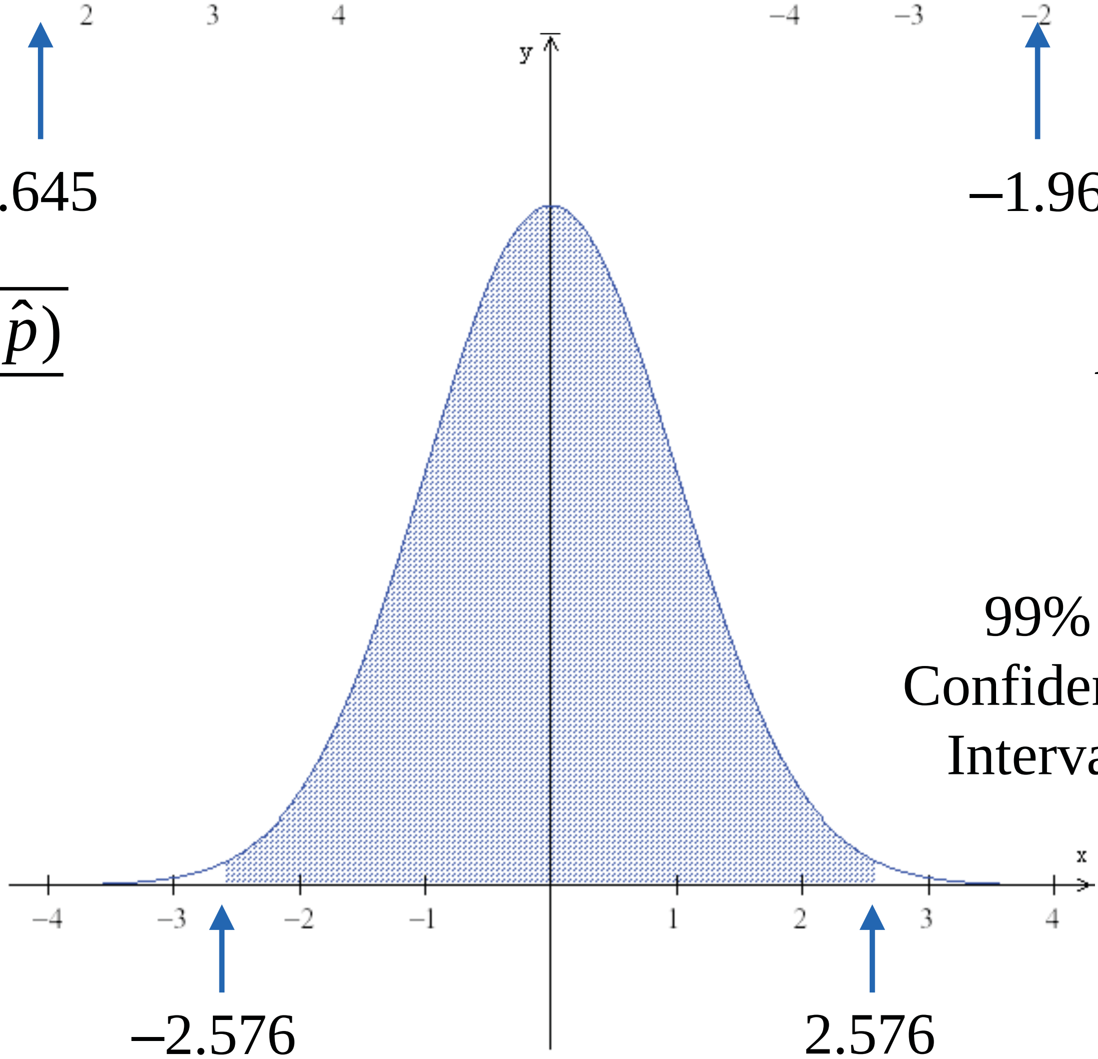
95%
Confidence
Interval



-1.96 $\qquad$ 1.96

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

99%
Confidence
Interval



-2.576 $\qquad$ 2.576

$$\hat{p} \pm 2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Margin of Error

General MOE Formula

$$\left(\underset{z\text{-score}}{\text{Critical Value}}\right)\left(\text{Standard Deviation}\right)$$

Standard Error

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

1 Sample Proportion MOE Formula

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

If $\hat{p}$ is unknown use $\hat{p} = 0.5$.

This will give us a conservative estimate for our sample size.

But why 0.5? Hint: Pre-Calc veterans can help here.

$$\hat{p}(1-\hat{p}) = \hat{p} - \hat{p}^2$$

Which is an upside down parabola that when graphed between 0 and 1 has it's max value at…?

$$\hat{p} = 0.5$$

Because we want our MOE to be large enough to contain the error estimation.

Assumptions for 1 Sample Proportion Confidence Intervals:

1. Random Sample or Sample Represents Population
2. $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$ ← Deal-breaker!
3. SSSRTP ← Allows us to sample without replacement

   Sample Sufficiently Small Relative to Population (10% rule)

Interpretation for 1 Sample Proportion Confidence Intervals

We are __% confident that p, the true proportion of ______, is between ___ and ___.

Interpretation for the Confidence <u>Level</u> of a 1 Sample Proportion Confidence Interval

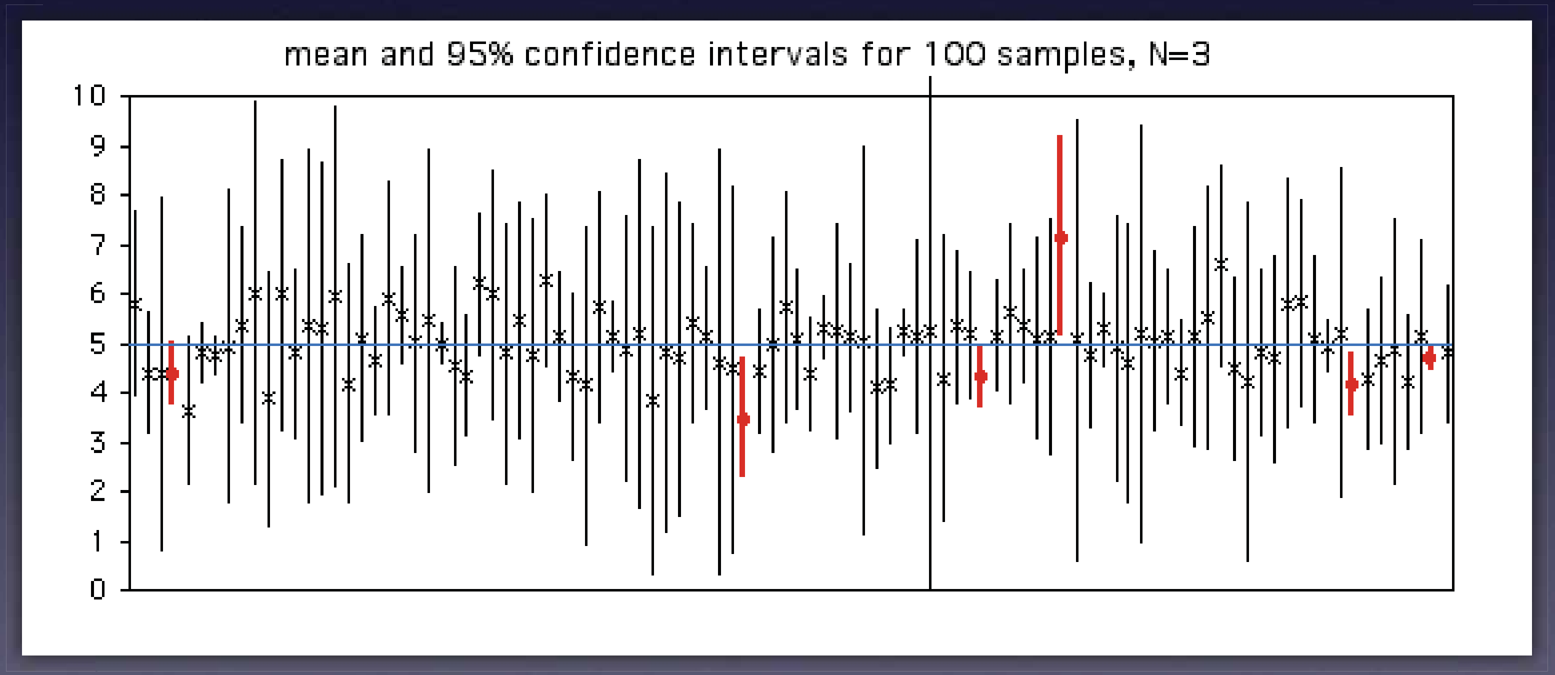
We used a method to construct this estimate that in the long run will successfully capture the true value of p ___% of the time

Interval vs. Level

A **confidence interval** gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter. We refer to this as the **confidence level**.

94 intervals were good
6 were bad



The higher the level, the wider the interval.

ALWAYS check your assumptions and interpret your interval, even you are not specifically asked to in the problem. Just do it. Seriously.

General Work Flow -

1. Assumptions (proportions from Unit 5)
2. Construction of (Confidence) Interval
3. Interpretation(s)

Try the examples and checkpoint questions in the notes

Raven, convinced that more of her classmates need to understand that Patrick Mahomes is the GOAT, takes a random survey of 50 SI students to see how many believe that Mahomes is, in fact, the GOAT. Her results are that 31 believe he is.

Construct a 90%, 95%, and 99% confidence interval for true proportion of Mahomes believers

Check assumptions

$50(.62) \geq 10$ ✓

$50(1 - 0.62) \geq 10$ ✓

50 is less than 10% of student body ✓

$$\hat{p} = \frac{31}{50} = 0.62 \qquad \sigma_{\hat{p}} = \sqrt{\frac{(.62)(.38)}{50}} = 0.0686$$

90% CI	95% CI	99% CI
$z = 1.645$	$z = 1.96$	$z = 2.576$
$0.62 \pm 1.645\sqrt{\frac{(.62)(.38)}{50}}$	$0.62 \pm 1.96\sqrt{\frac{(.62)(.38)}{50}}$	$0.62 \pm 2.576\sqrt{\frac{(.62)(.38)}{50}}$
(0.507, 0.733)	(0.485, 0.755)	(0.443, 0.797)

Interpretation for 1 Sample Proportion Confidence Interval
We are 90% confident that *p*, the true proportion of Mahomes believers, is between 0.507 and 0.733 (or between 50.7% and 73.3%)

90%

Confidence <u>Level</u>
We used a method to construct this estimate that in the long run will successfully capture the true value of *p* 90% of the time

Interpretation for 1 Sample Proportion Confidence Interval
We are 95% confident that *p*, the true proportion of Mahomes believers, is between 0.485 and 0.755 (or between 48.5% and 75.5%)

95%

Confidence <u>Level</u>
We used a method to construct this estimate that in the long run will successfully capture the true value of *p* 95% of the time

Interpretation for 1 Sample Proportion Confidence Interval
We are 99% confident that *p*, the true proportion of Mahomes believers, is between 0.443 and 0.797 (or between 44.3% and 79.7%)

99%

Confidence <u>Level</u>
We used a method to construct this estimate that in the long run will successfully capture the true value of *p* 99% of the time

Raven, convinced that more of her classmates need to understand that Patrick Mahomes is the GOAT, takes a random survey of 50 SI students to see how many believe that Mahomes is, in fact, the GOAT. Her results are that 31 believe he is.

Those were some pretty wide intervals. How do you supposed we could reduce them?

How about a sample size of 140?

$$\hat{p} = \frac{87}{140} = 0.621 \quad \sigma_{\hat{p}} = \sqrt{\frac{(.621)(.379)}{140}} = 0.041$$

Check assumptions

$140(.62) \geq 10$ ✓

$140(1 - 0.62) \geq 10$ ✓

140 is barely less than 10% of student body ✓

95% CI with sample size = 50

$z = 1.96$

$$0.62 \pm 1.96\sqrt{\frac{(.62)(.38)}{50}}$$

(0.485, 0.755)

95% CI with sample size = 140

$z = 1.96$

$$0.62 \pm 1.96\sqrt{\frac{(.62)(.38)}{140}}$$

(0.541, 0.702)

Raven, convinced that more of her classmates need to understand that Patrick Mahomes is the GOAT, takes a random survey of 50 SI students to see how many believe that Mahomes is, in fact, the GOAT. Her results are that 31 believe he is.

95% CI with sample size = 50

$$0.62 \pm 1.96\sqrt{\frac{(.62)(.38)}{50}}$$

(0.485, 0.755)

95% CI with sample size = 140

$$0.62 \pm 1.96\sqrt{\frac{(.62)(.38)}{140}}$$

(0.541, 0.702)