Tuesday's (4/26/22) class will be on zoom.

- There will not be a Top Hat activity for Tuesday's class
- You can attend 8:30 or 12 pm
- Recordings from both will be posted on canvas

Week 4

Confidence Intervals for a Proportion

ST 314
Introduction to Statistics for Engineers

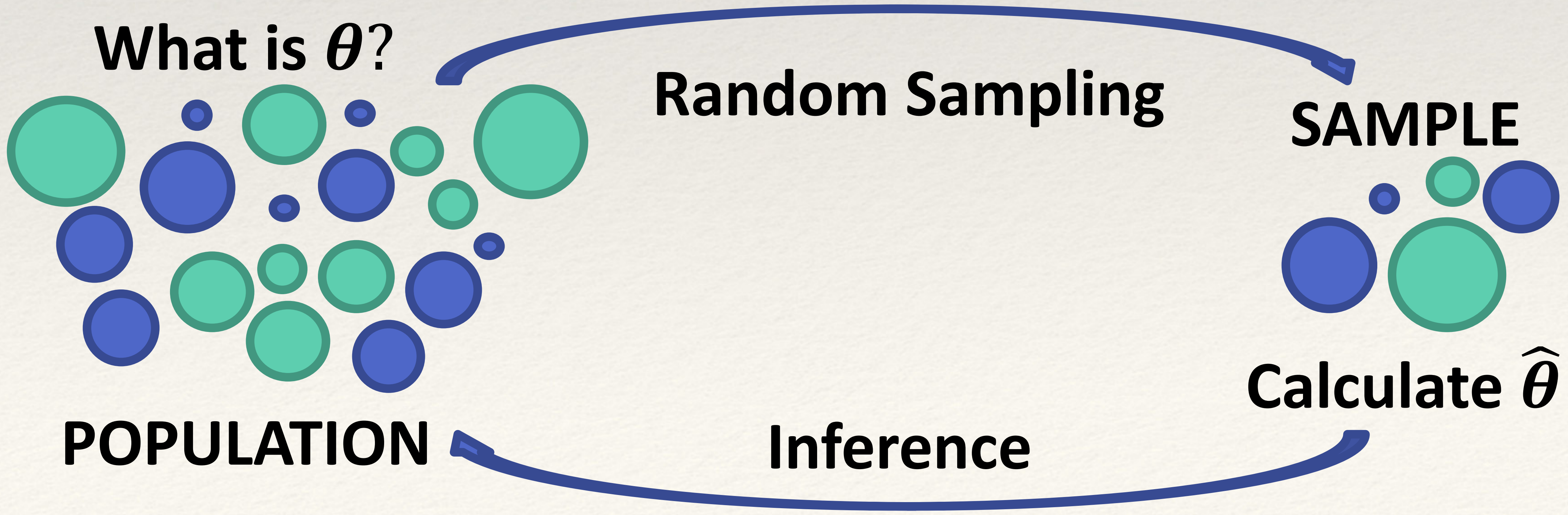
Oregon State University

Estimation

- A **point estimate** is the single value considered to be the "best guess" for the parameter } sample statistic

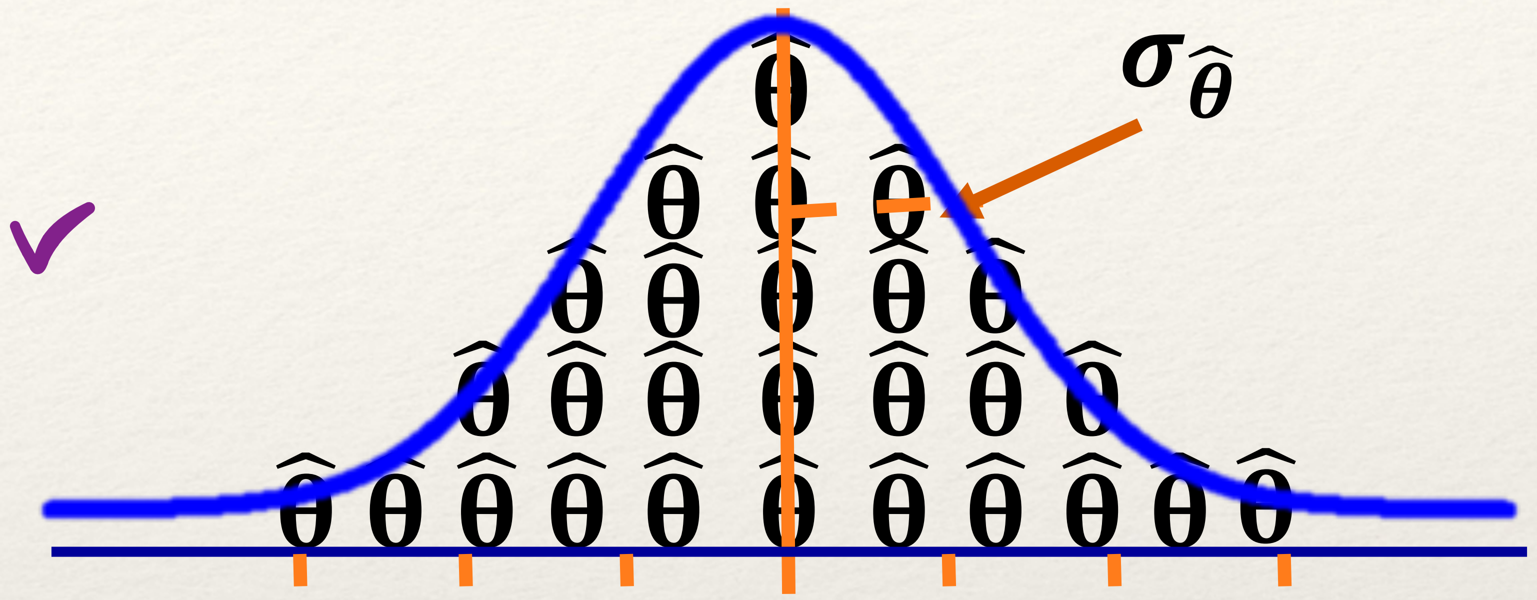
- An **unbiased** point estimate is one in which the expected value is equal to the parameter: $E(\hat{\theta}) = \theta$.

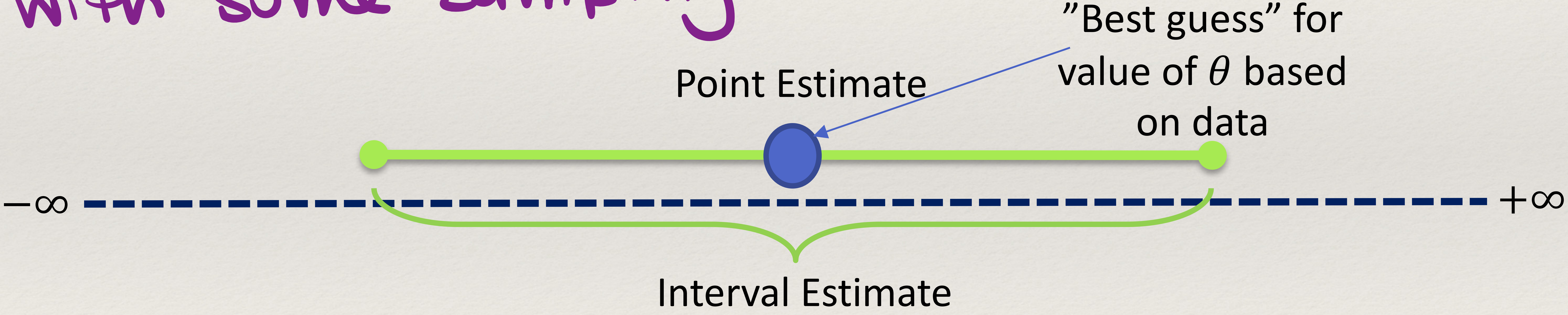
- It is desirable to use estimates that are unbiased and have a small standard error.



Estimation and Sampling Variability

A point estimate is rarely equal to the true parameter and is usually associated with some sampling variability.





A **confidence interval** can increase the likelihood an estimation procedure will capture the population parameter. An interval estimate provides a range of reasonable values for the population parameter where the point estimate is the center of the interval.

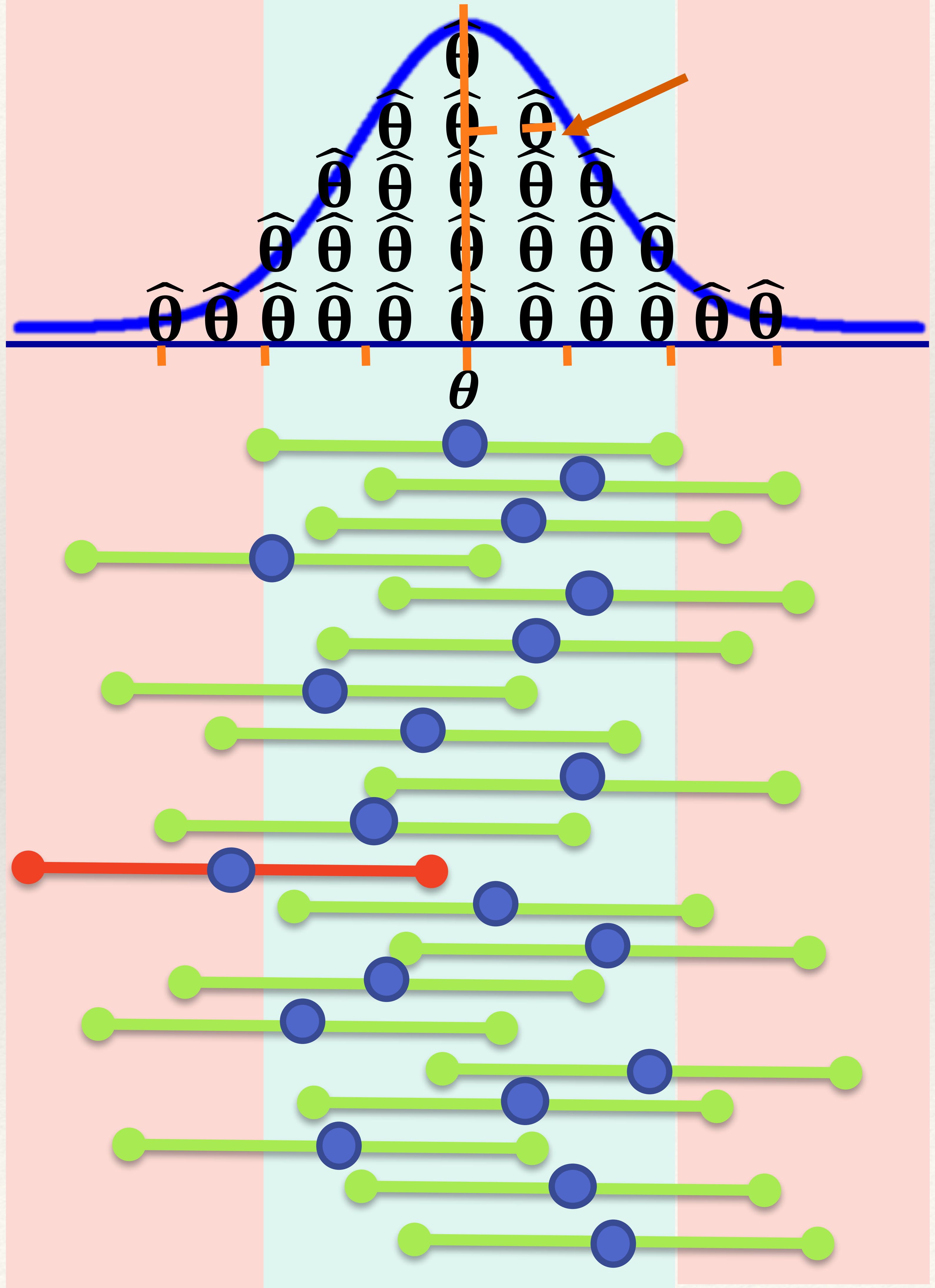
Constructing a 95% Confidence Interval

- If the sampling distribution of $\hat{\theta}$ is normal, then 95% of all possible estimates will be within 1.96 standard errors of the population parameter.

- If a point estimate within 1.96 standard errors is selected, then the interval $\hat{\theta} \pm 1.96\,\sigma_{\hat{\theta}}$ will capture the parameter, θ.

- We can be 95% confident the parameter θ is captured by the above interval estimate.

- 5% of intervals won't capture θ. Since we don't know the value of θ, we won't know if the interval we constructed is one of the "good ones".

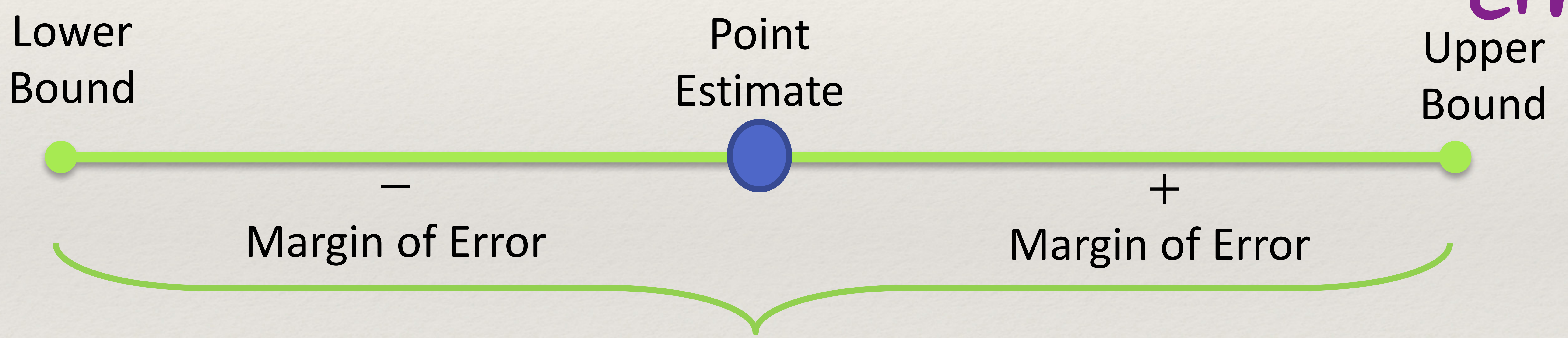


Constructing a Confidence Interval

- All confidence intervals have the same general form:

  point estimate $\pm$ margin of error

- The **margin of error** is the product of a critical value and the standard error.



For a $100(1-\alpha)\,\%$ confidence interval, the critical value is the $1-\dfrac{\alpha}{2}$ percentile in the standardized sampling distribution.

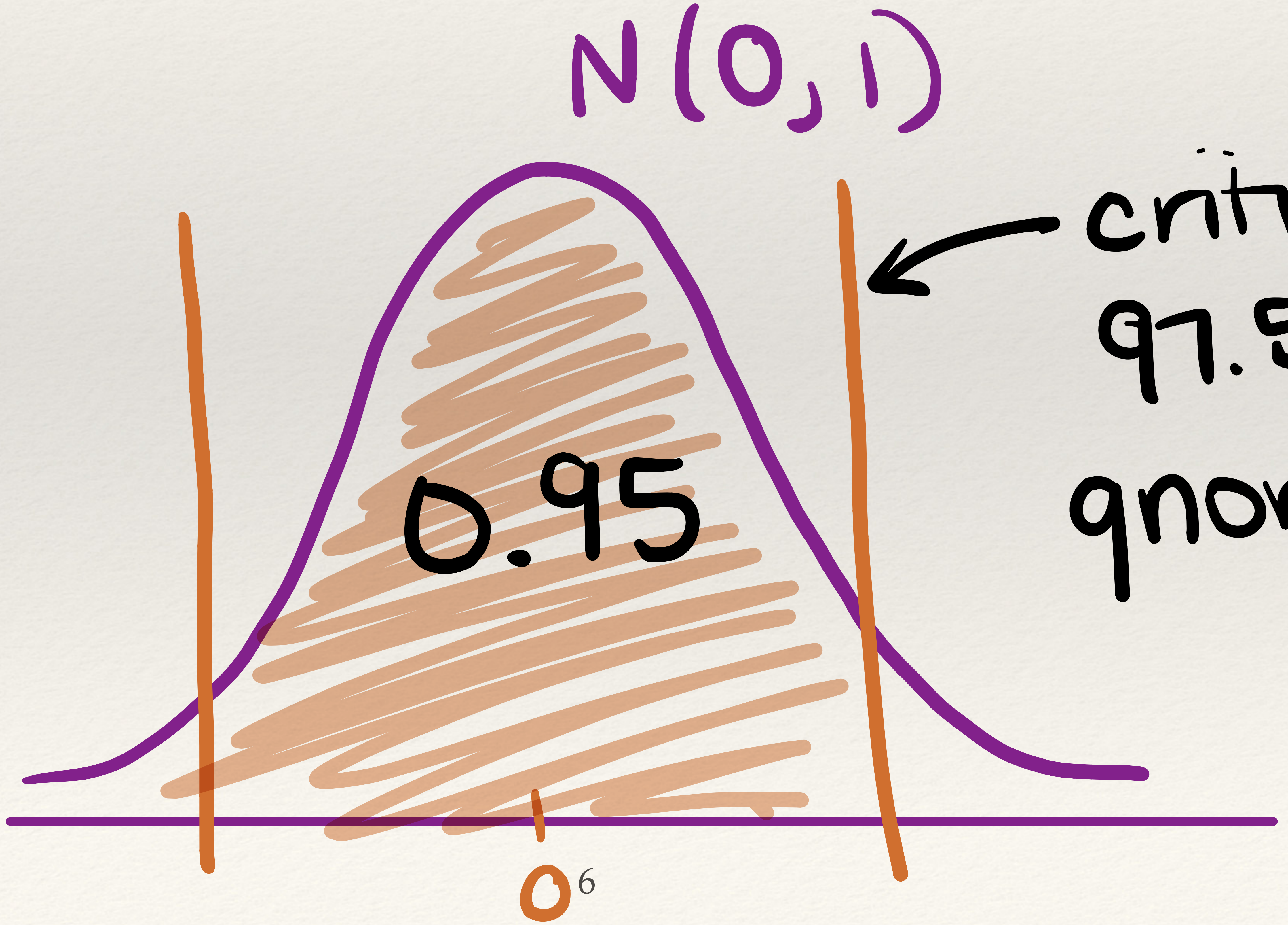
Constructing a Confidence Interval

- For a $100(1 - \alpha)\,\%$ confidence interval, the critical value is the $\underline{\;1-\frac{\alpha}{2}\;}$ percentile in the standardized sampling distribution.

- $\underline{\;\alpha\;}$ is the error of the interval and is determined by the desired level of confidence.

95% CI

$\alpha = 0.05$



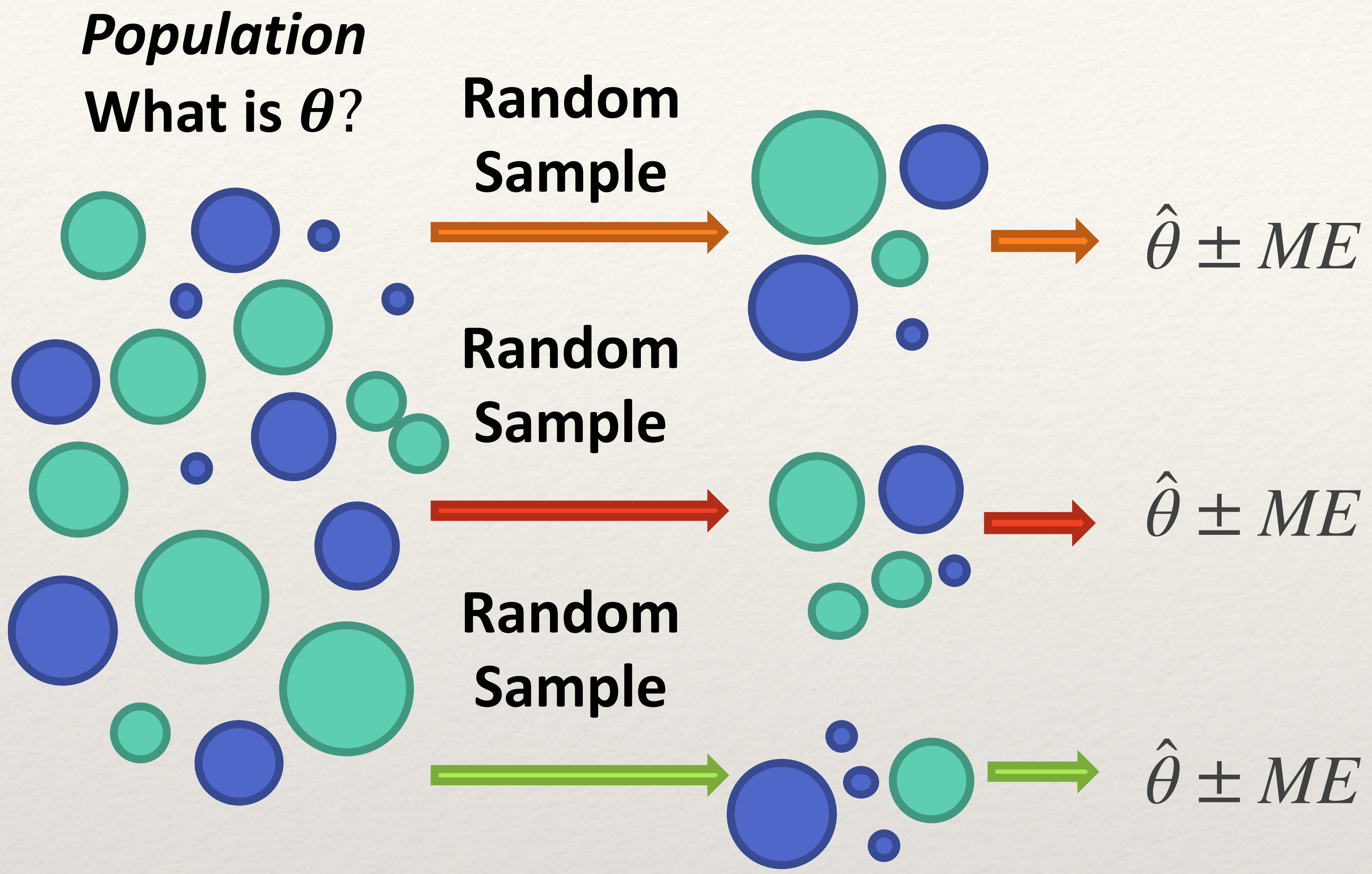
critical value

97.5th percentile

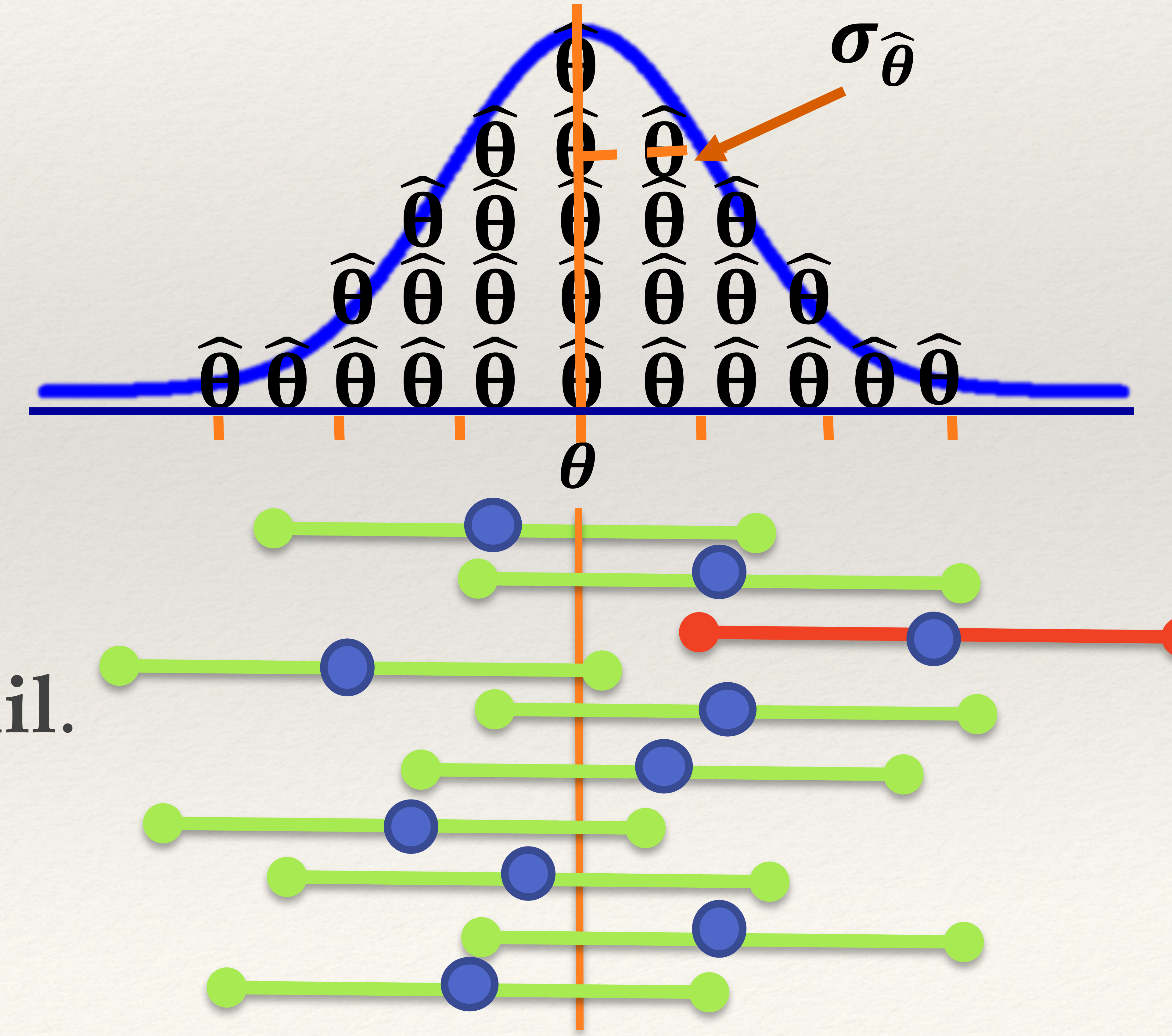
qnorm(0.975)

=1.96

"Confidence" in Confidence Intervals



The term "confidence" comes from theory based on repeated sampling.



If a **90% CI** is calculated from every random sample, then on average 9 **out of 10 intervals will capture the parameter**, where as 10% **of all intervals will fail.**

Therefore, we are 90% confident our interval will capture the parameter.

How to Describe a Confidence Interval

confidence level

population parameter

The ________% confidence interval estimates the ______________ to be between ______ and ______, with a point estimate of ______.

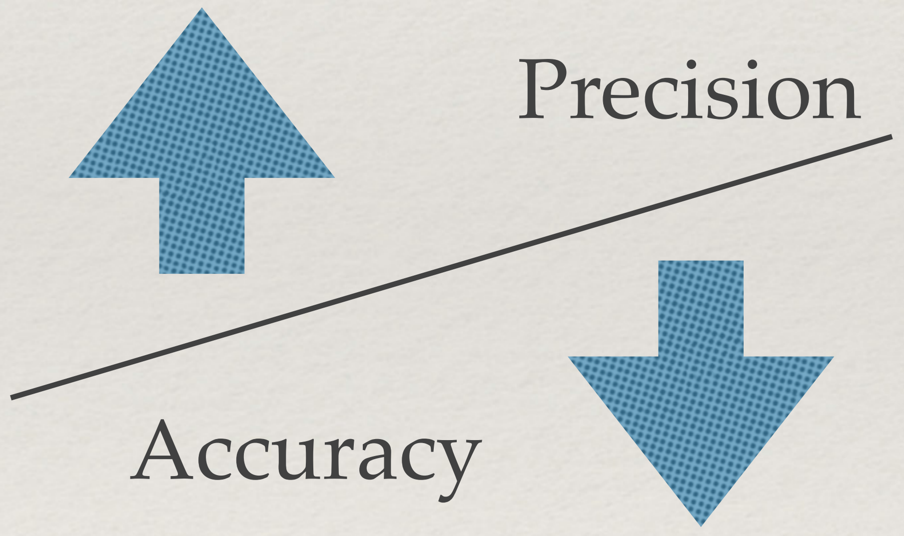
lower bound

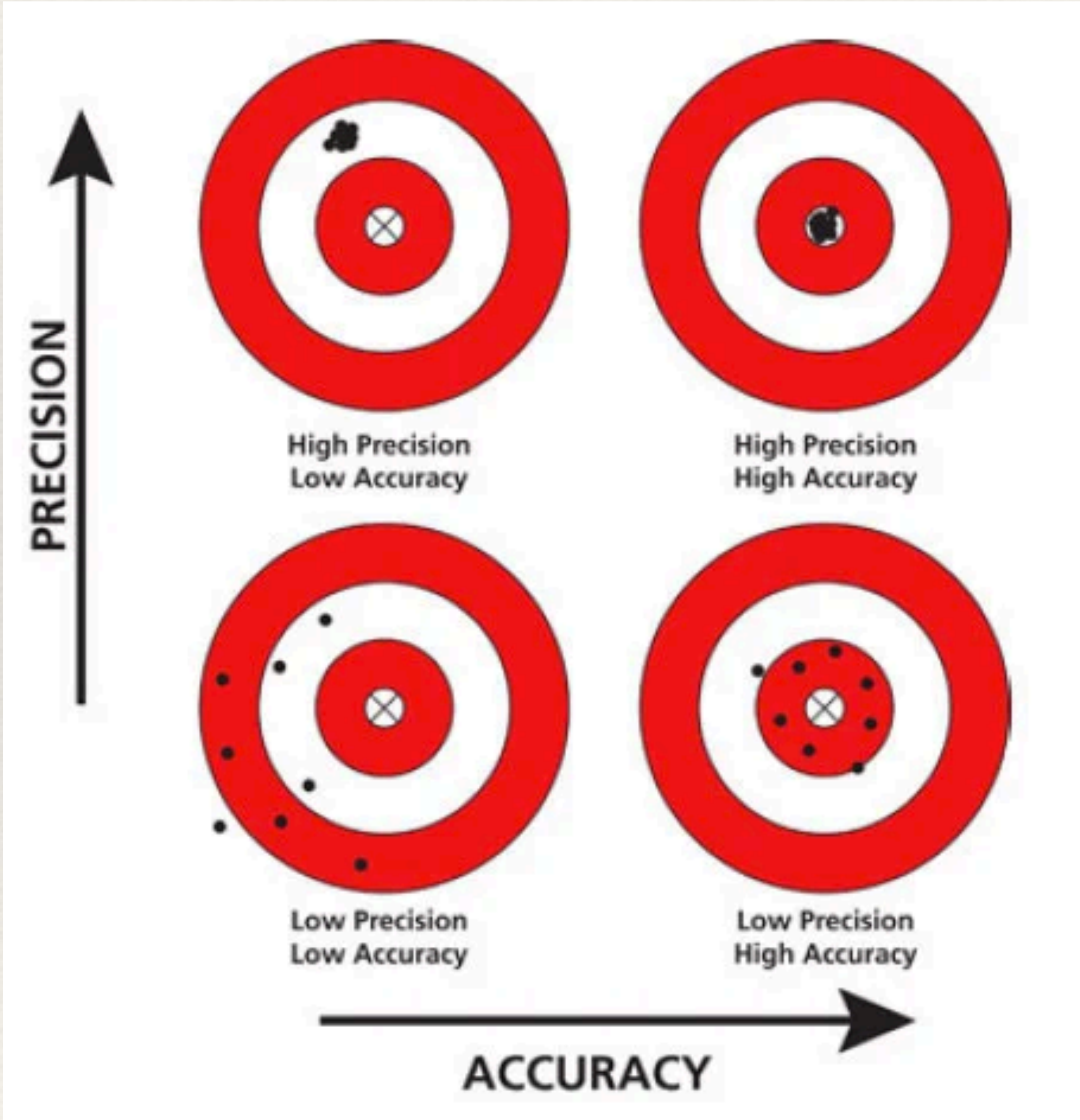
upper bound

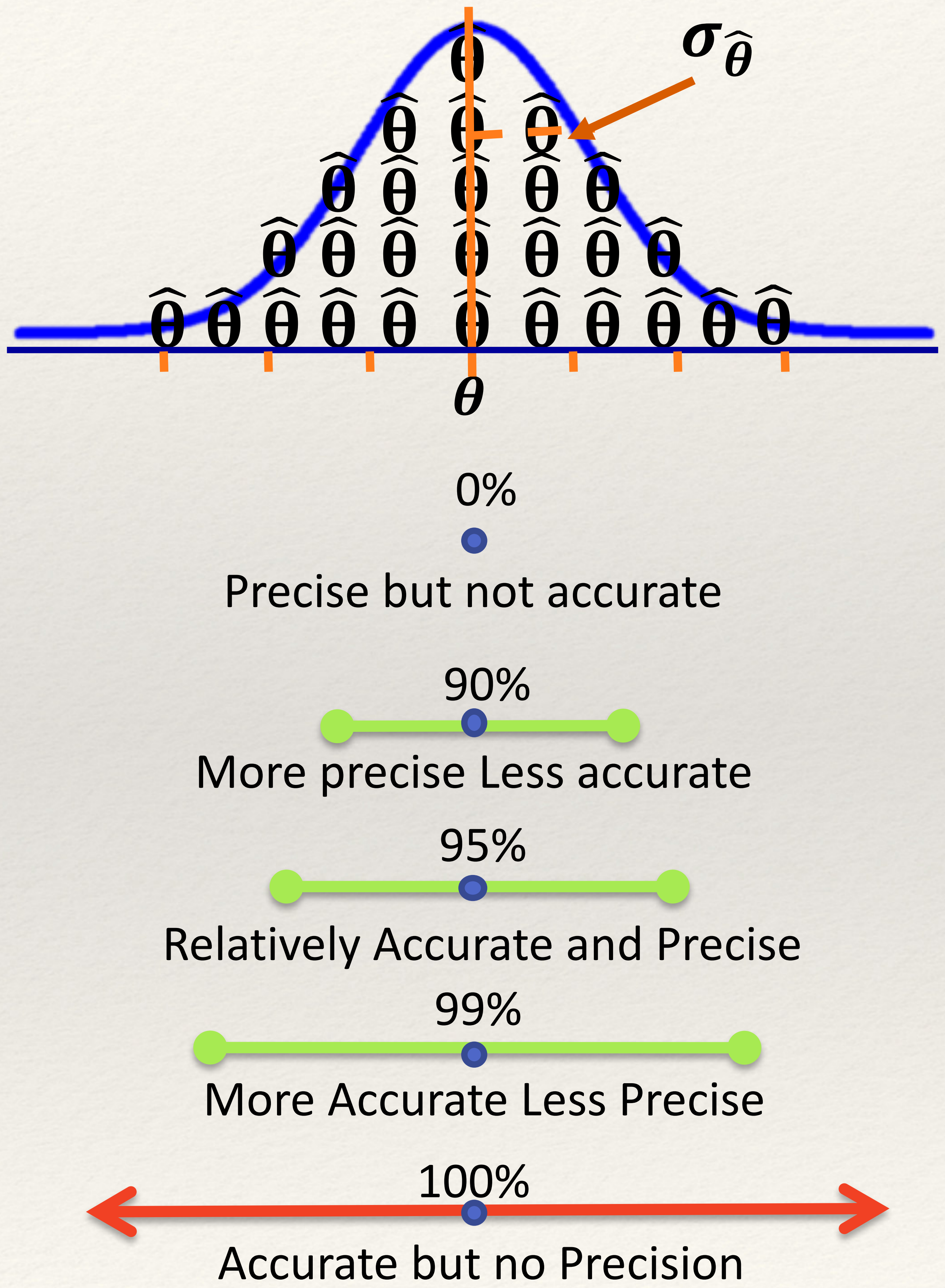
point estimate

Choosing a Confidence Level

The confidence level is chosen based on whether precision or accuracy is more desirable.

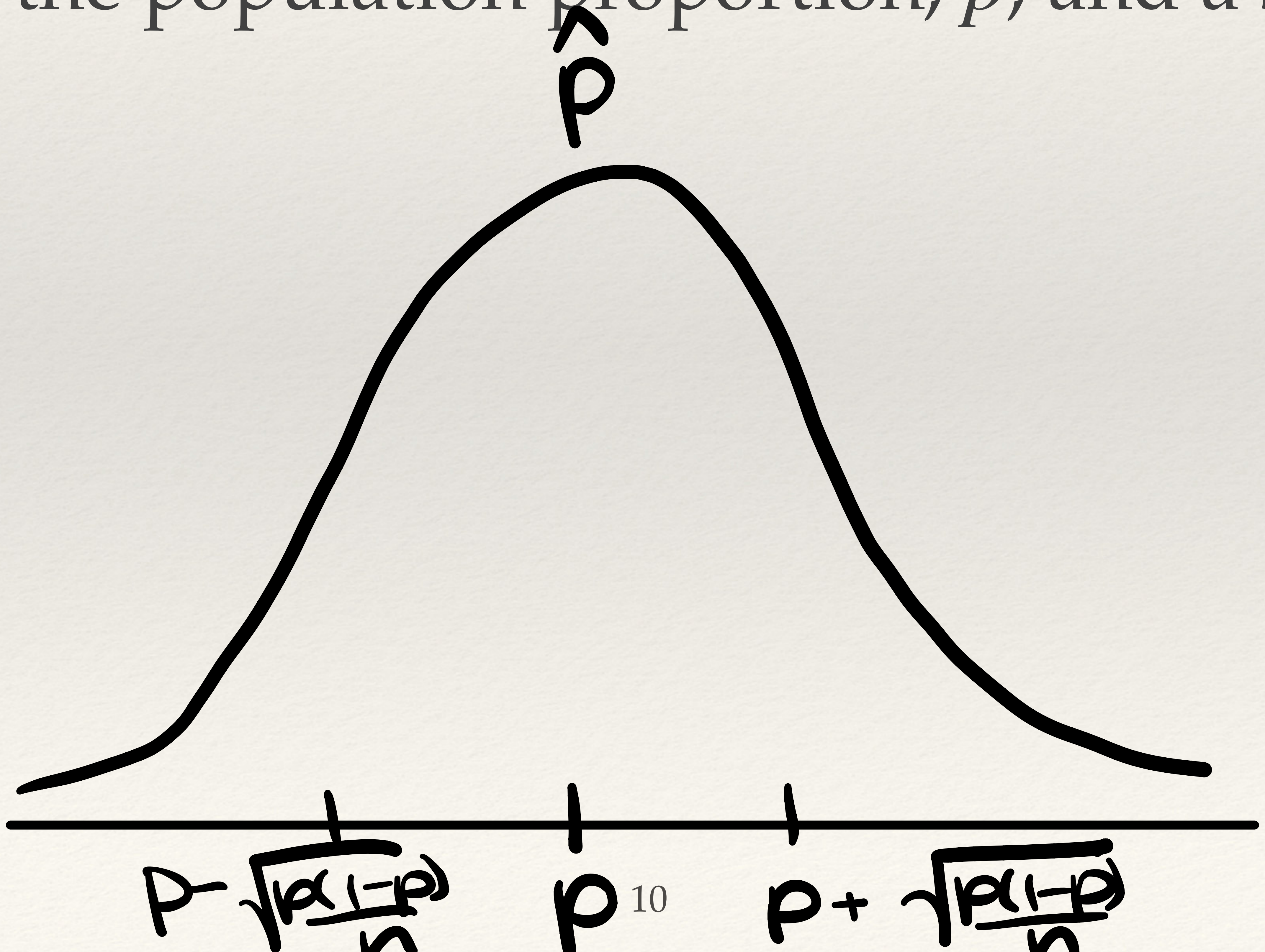






Central Limit Theorem for a Proportion

Recall that for a sample of size n, if n is sufficiently large, then the **Central Limit Theorem** states that $\hat{p}$ follows an approximately Normal distribution with a mean equal to the population proportion, p, and a standard deviation of $\sqrt{\dfrac{p(1-p)}{n}}$.



Confidence interval for p

- When to use:

want to estimate the population proportion p from a single population

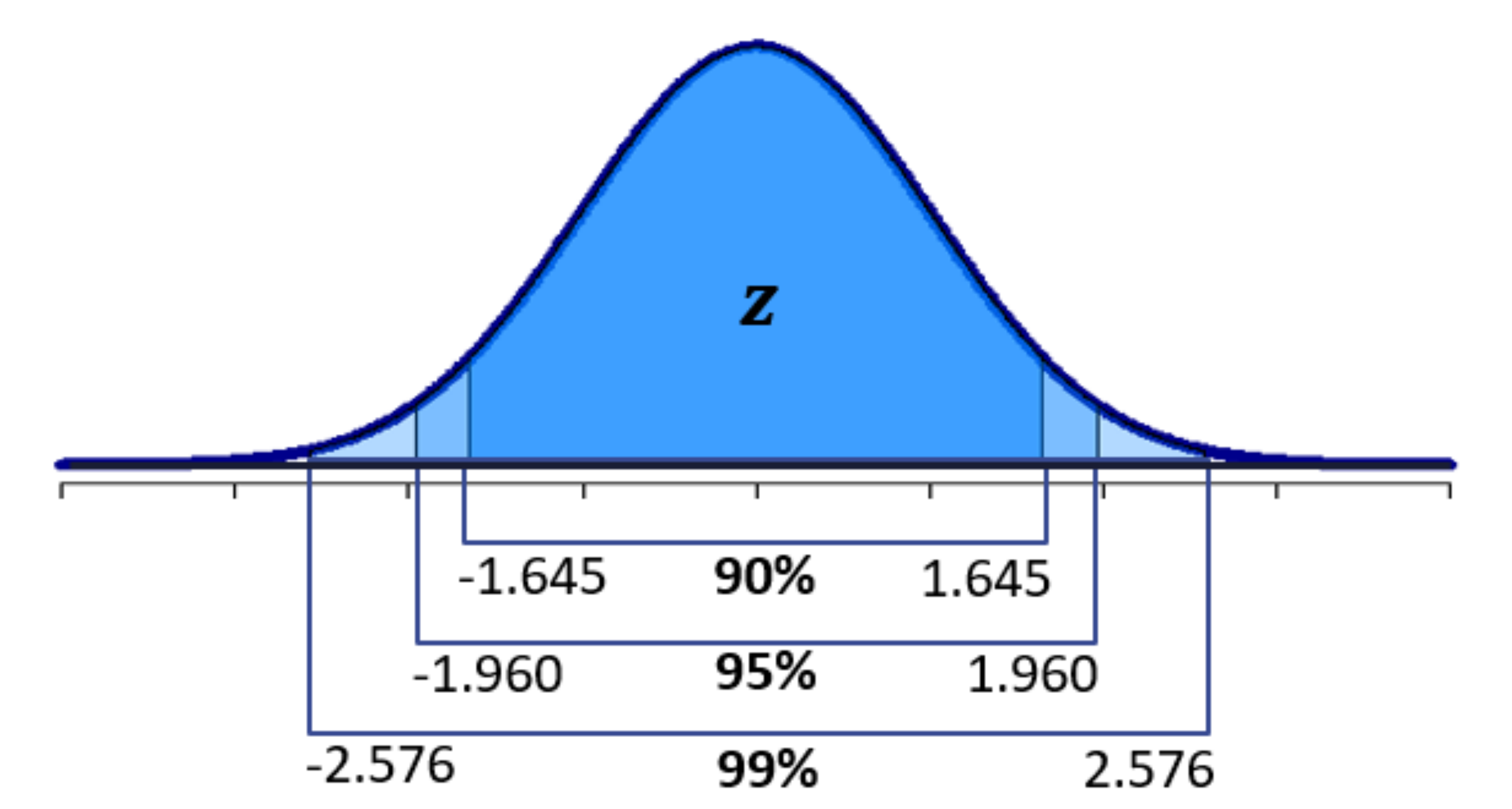
- The confidence interval for a population proportion is:

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- The interval above relies on the Central Limit Theorem. Therefore, we need a sufficiently large sample size. Since we don't know *p*, we need to check the following: $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

Critical Values for *z* Confidence Intervals

Confidence Level	$z_{1-\frac{\alpha}{2}}$ Critical values
90%	1.645
95%	1.96
99%	2.576



Confidence Interval Example

A 2013 survey of 434 adults found that 52% of U.S. adult Twitter users get at some news on Twitter. Construct a 99% confidence interval for the proportion of U.S. adult Twitter users who get some news on Twitter and interpret the interval in context.

$\hat{p} = 0.52 \quad n = 434$

sample size conditions: $n\hat{p} = 434(0.52) = 225.68 \geq 10$ ✓

$n(1-\hat{p}) = 434(0.48) = 208.32 \geq 10$ ✓

$$\hat{p} \pm \underbrace{z_{1-\frac{\alpha}{2}}}_{2.576} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.52 \pm 2.576 \sqrt{\frac{0.52(0.48)}{434}} = (0.458, 0.582)$$

Confidence Interval Example

A 2013 survey of 434 adults found that 52% of U.S. adult Twitter users get at some news on Twitter. Construct a 99% confidence interval for the proportion of U.S. adult Twitter users who get some news on Twitter and interpret the interval in context.

We are 99% confident that the proportion of adult twitter users in US that get news from twitter is between 0.458 and 0.582 with a point estimate of 0.52.