Introduction to Confidence Intervals

"A 95% confidence for the mean μ is given by (-2.14,3.07)."

This does **NOT** mean:

- You are 95% "confident" that the true mean μ is between -2.14 and 3.07.
- The true mean μ is between -2.14 and 3.07 with probability 0.95.

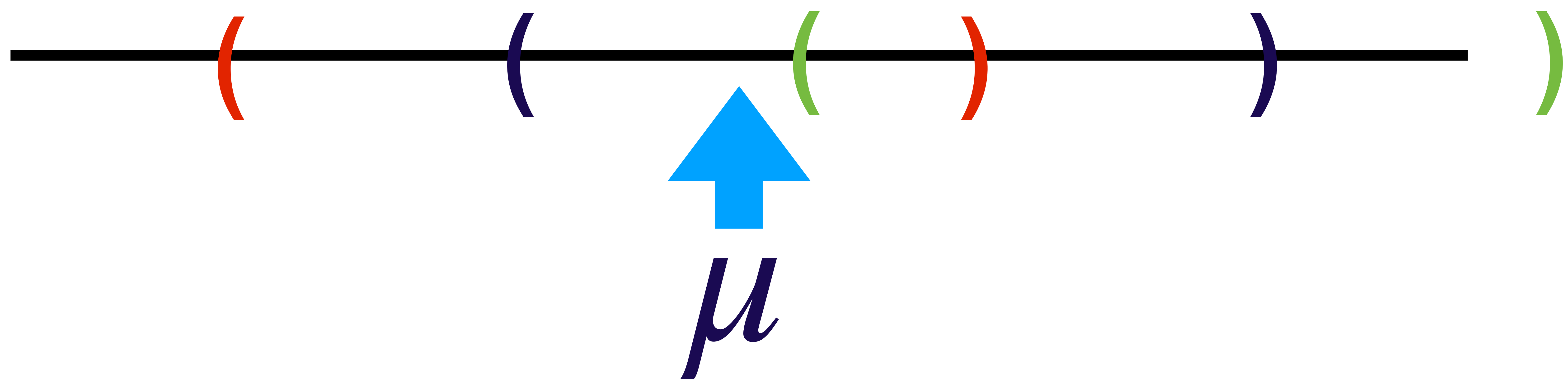
Introduction to Confidence Intervals

"A 95% confidence for the mean μ is given by (-2.14,3.07)."

The randomness is in the sampling.

- Collect your sample.
- Estimate the parameter.
- Return a confidence interval.

If you did this again, you would not get the same results!



Multiple samples give
multiple confidence intervals.

$\boxed{\text{95\% of them will correctly capture } \mu.}$

From the Ground Up:

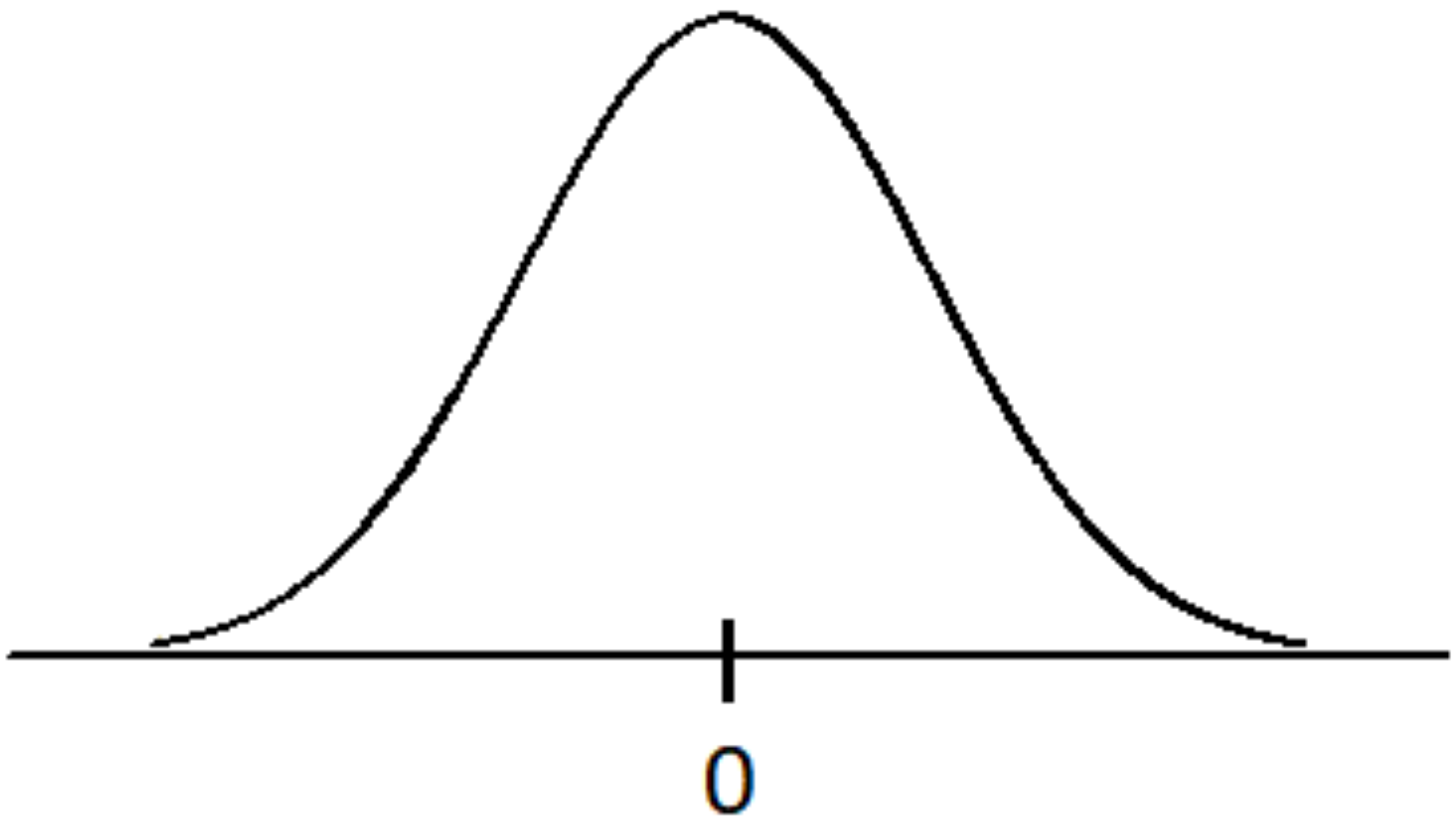
Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the normal distribution with mean μ and variance σ^2.

Assume that σ^2 is known.

- $\overline{X}$ is an estimator of μ
- $\overline{X}$ is $N(\mu, \sigma^2/n)$.

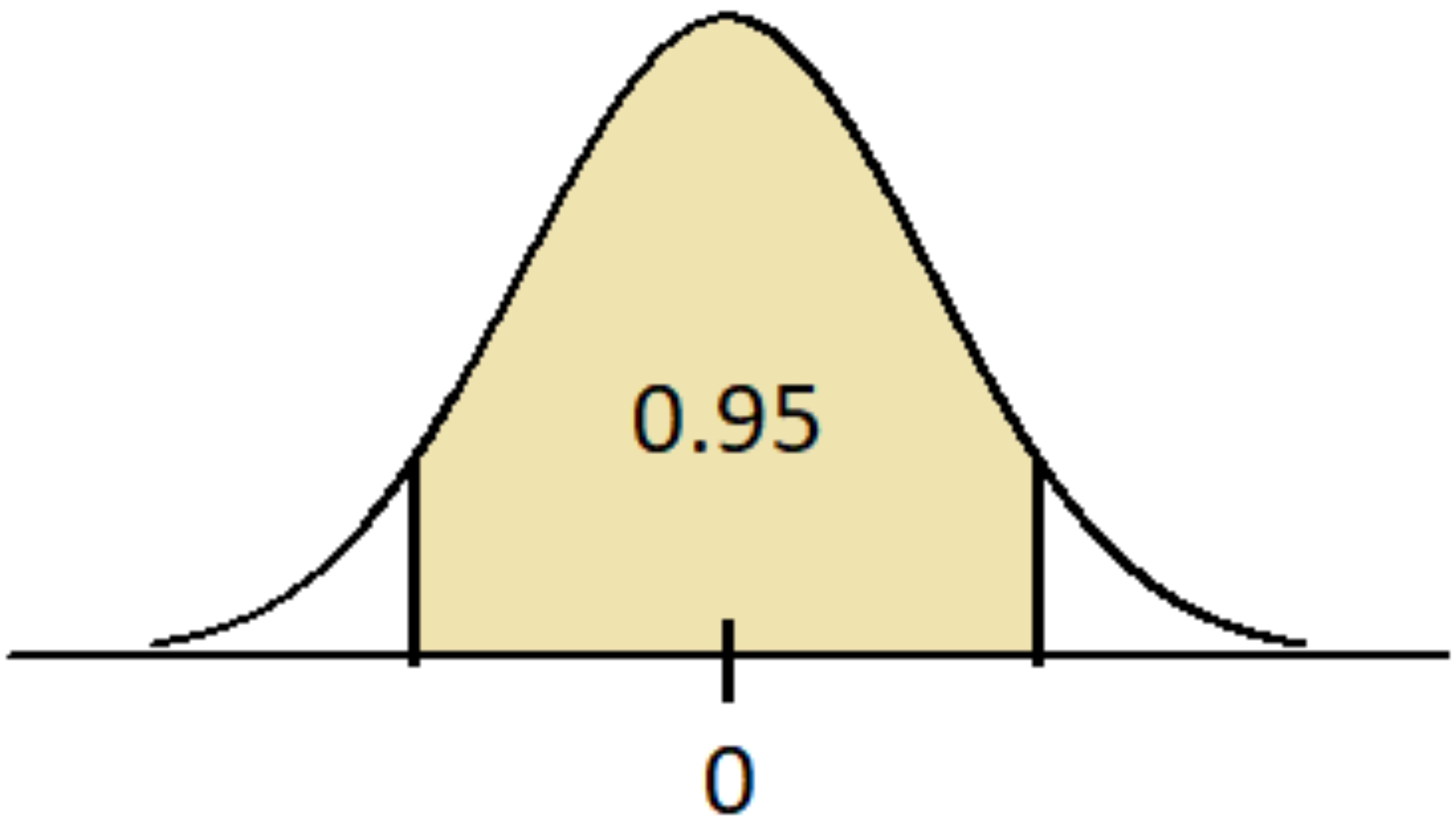
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Suppose that $Z \sim N(0, 1)$.



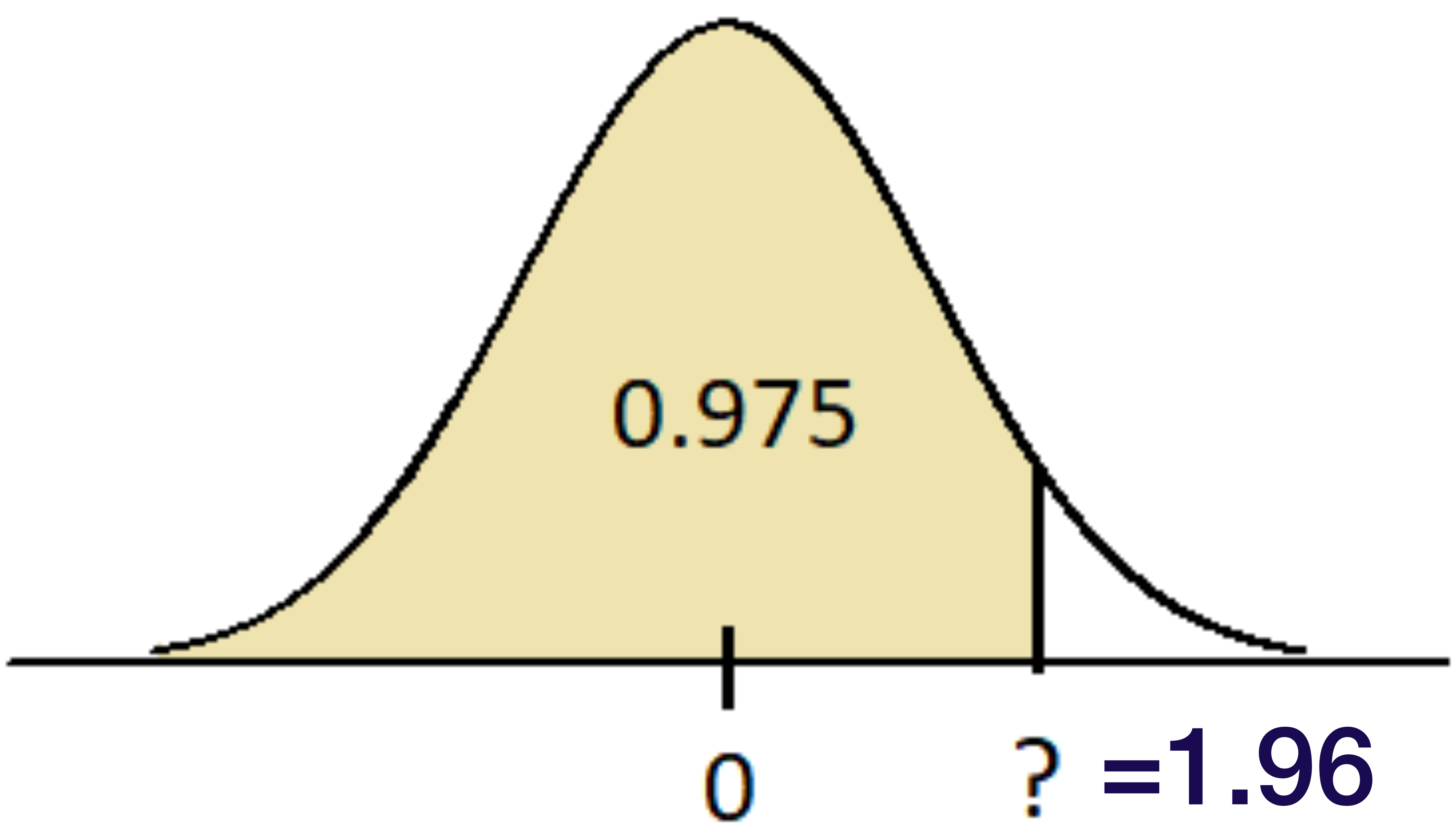
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Suppose that $Z \sim N(0, 1)$.



$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Suppose that $Z \sim N(0,1)$.



In R: qnorm(0.975)

$$P(-1.96 < Z < 1.96) = 0.95$$

$$\Downarrow$$

$$P\left(-1.96 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

$$\Downarrow$$

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}1.96\right) = 0.95$$

A 95% confidence interval for the mean μ of a normal distribution is given by

$$\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + \frac{\sigma}{\sqrt{n}}\right)$$

This can be written as

$$\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

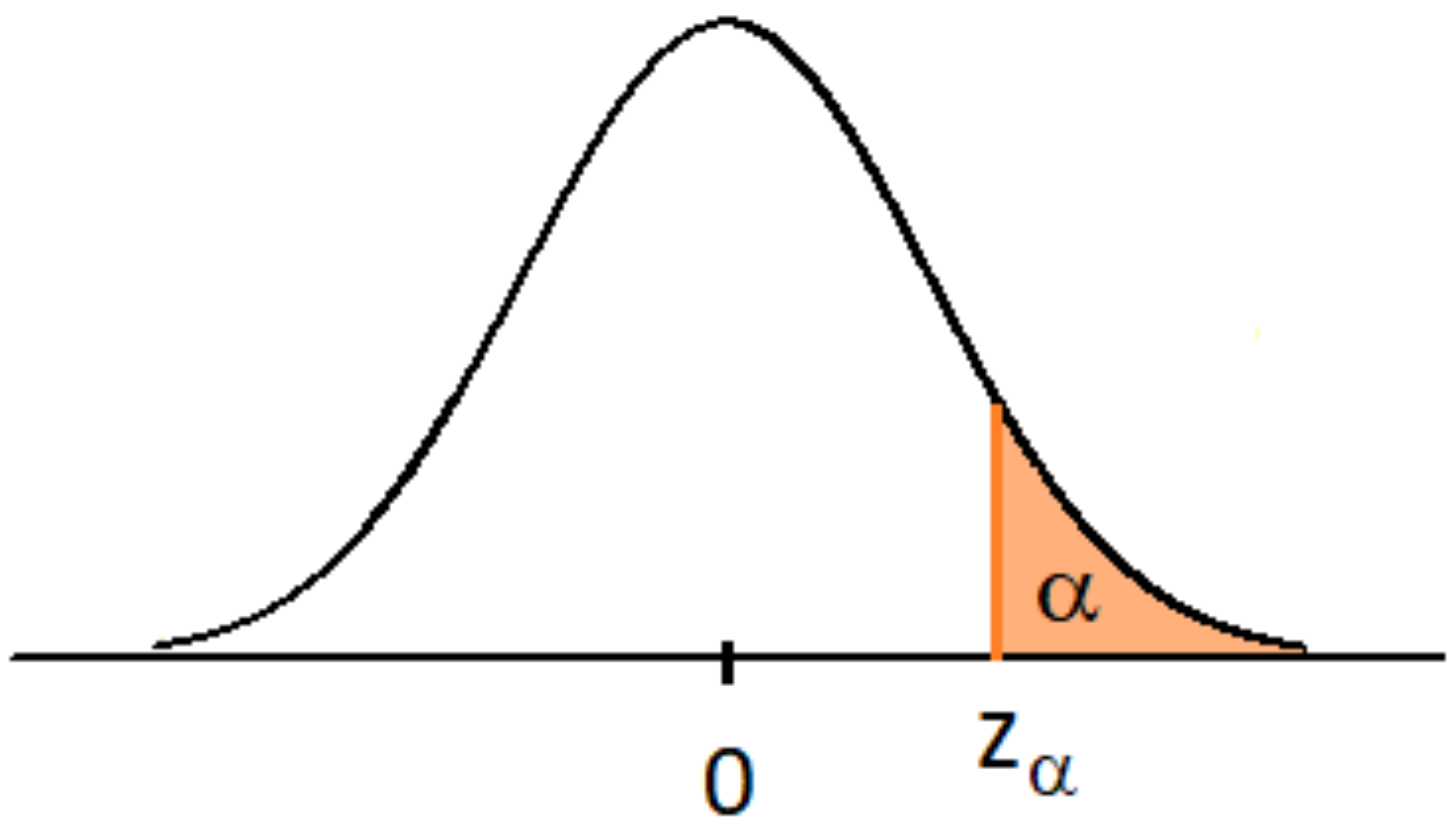
The 1.96 is called a critical value.

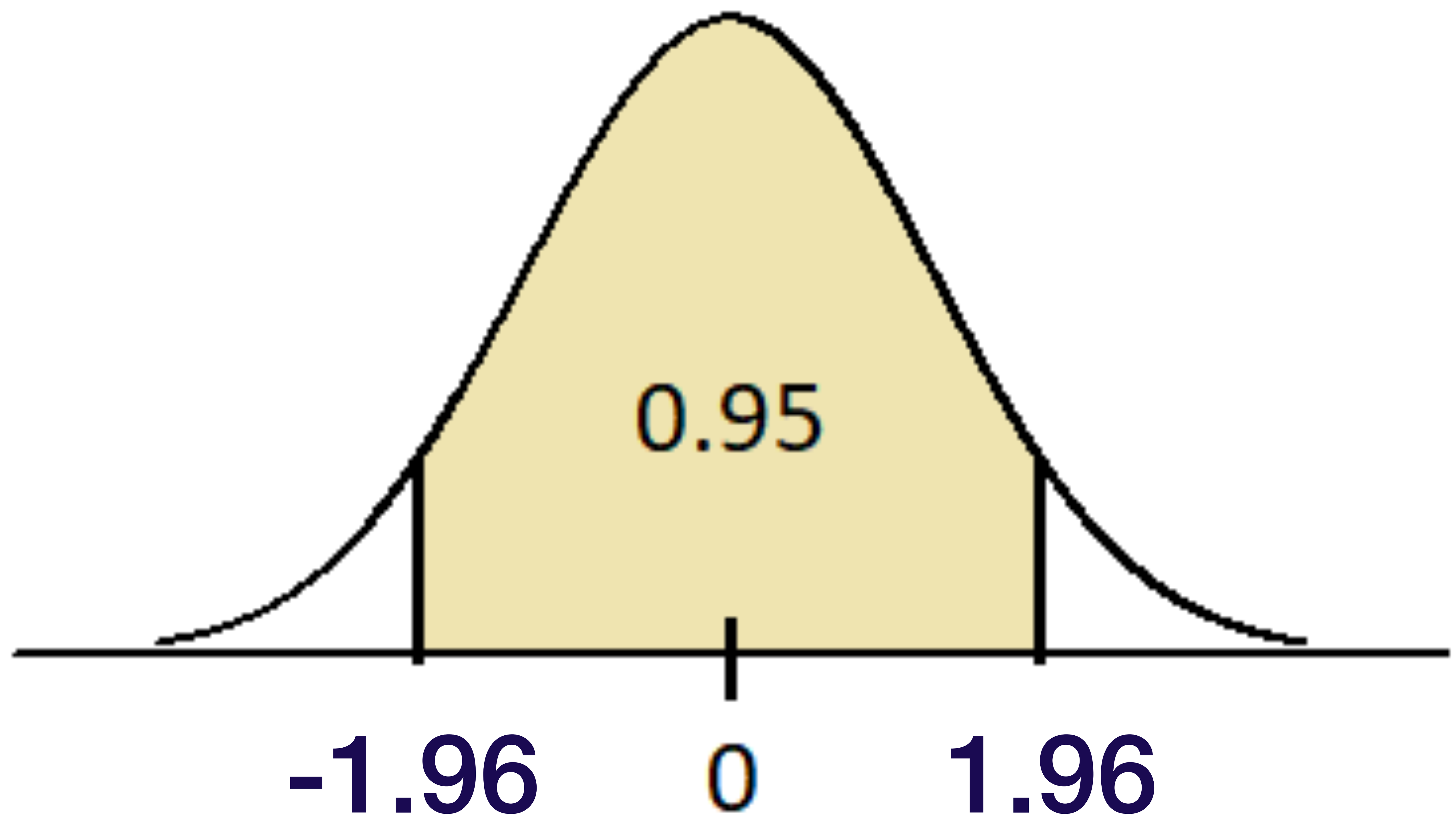
In general, a critical value is a number that cuts off a specified area under a pdf.

Notation:

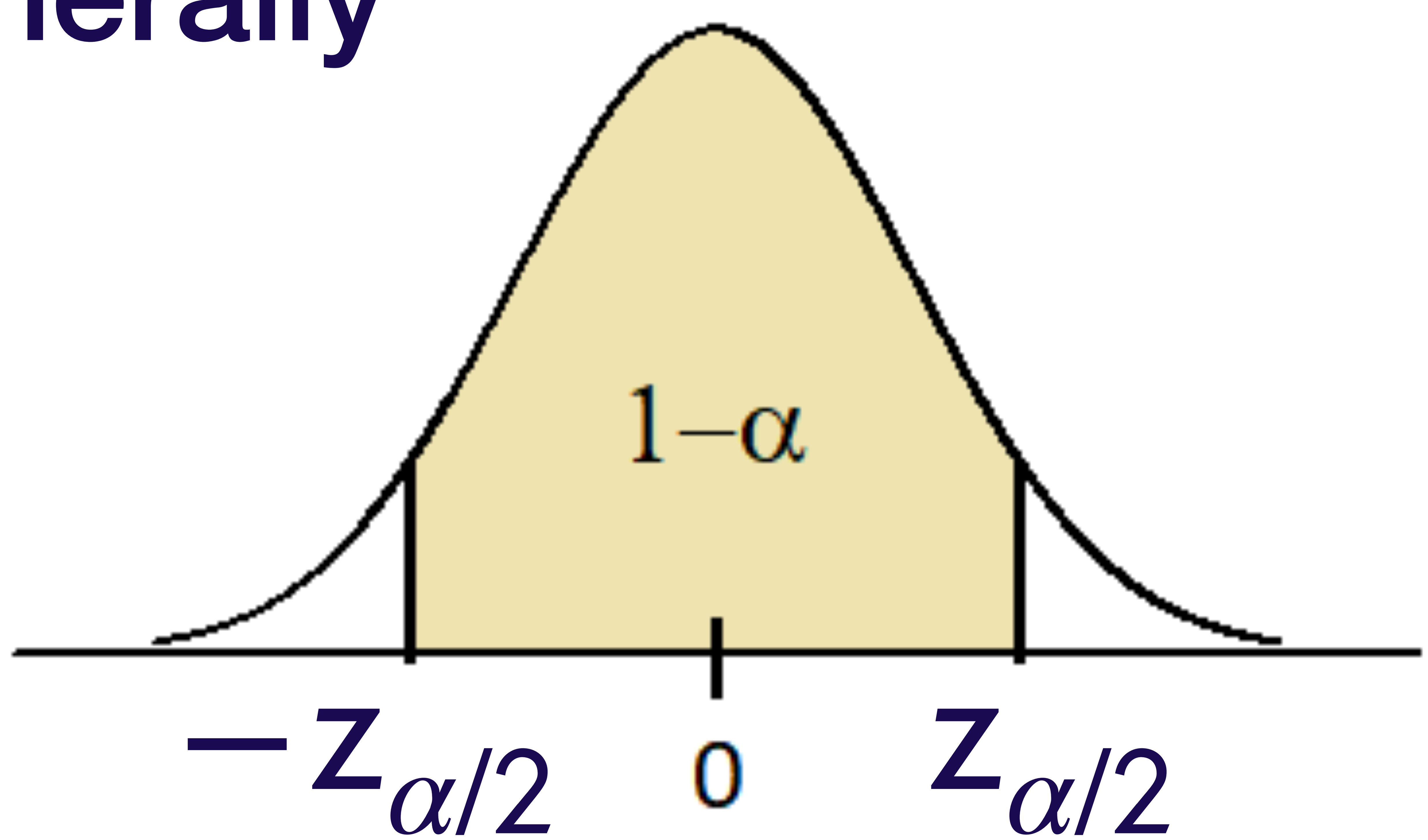
Let $Z \sim N(0,1)$.

Let z_α be the # that cuts off area α to the right.





More generally



$$-z_{\alpha/2} = z_{1-\alpha/2}$$

Summary

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the normal distribution with mean μ and variance σ^2.

A $100(1-\alpha)\%$ confidence interval for μ is given by

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Note

Everything we did was based on the fact that $\overline{X}$ has a normal distribution.

This came from the fact that the sample came from a normal distribution.

CLT: For a more general distribution

$\overline{X}$ has roughly a normal distribution for large samples $(n>30)$

An "Important Thing"

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the any distribution with mean μ and variance $\sigma^2 < \infty$.

For large n, an approximate $100(1-\alpha)\%$ confidence interval for μ is given by

$$\boxed{\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

Example

A supermarket chain is considering adding more organic produce to its offerings in a certain region of the country. They hired an external marketing company which collected data for them.

Based on a random sample of 200 customers from the region, they observed that the average amount spent on organic produce, per person and per month, was $36.

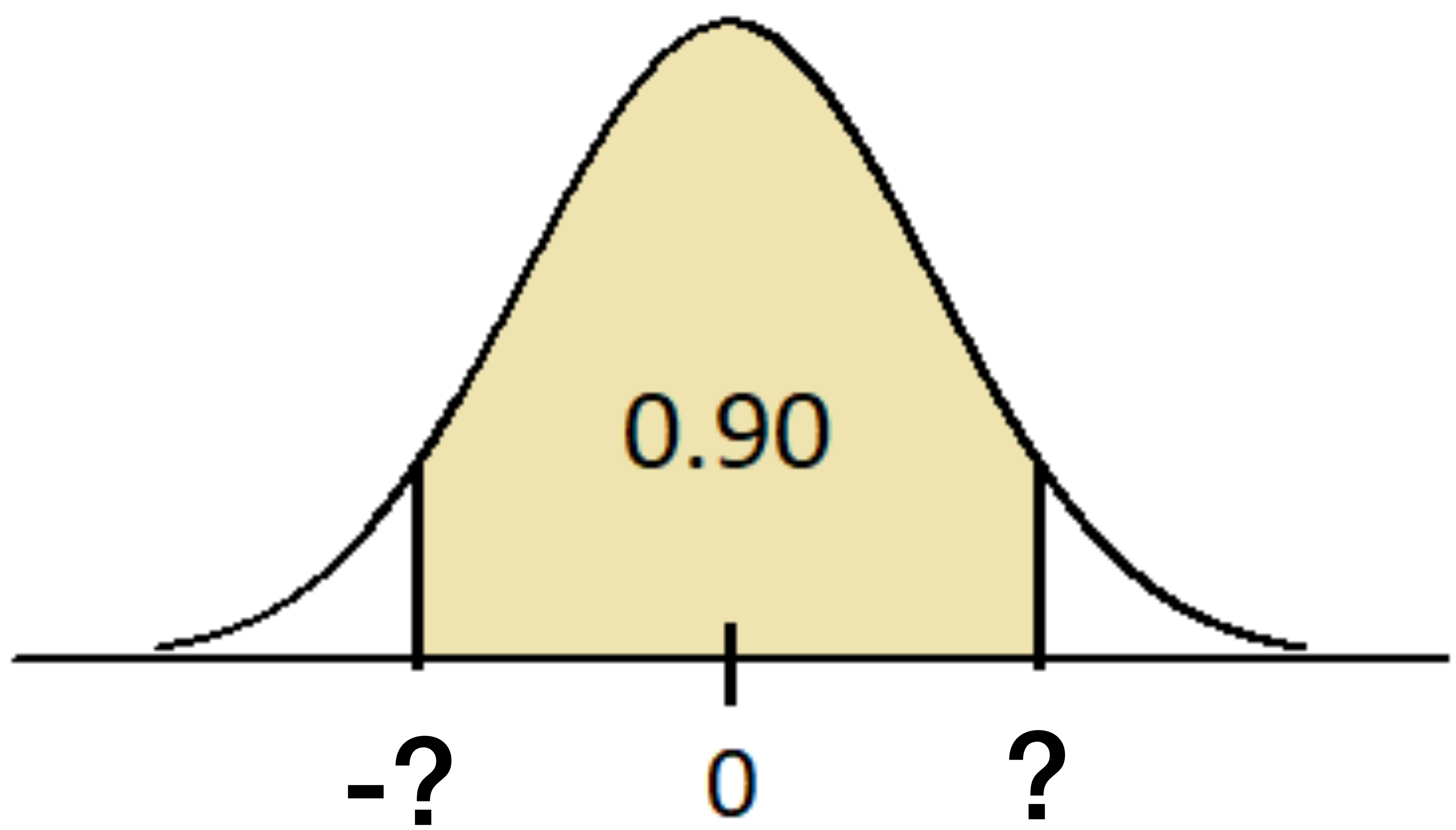
Based on past studies, it is believed that the variance of the amount spent on produce, organic or not, is 5 dollars.

Example

Find a 90% confidence interval for the true average dollar amount that all customers in the region spent on organic produce each month.

We have

- $n = 200$
- $\bar{x} = 36$
- $\sigma^2 = 5$

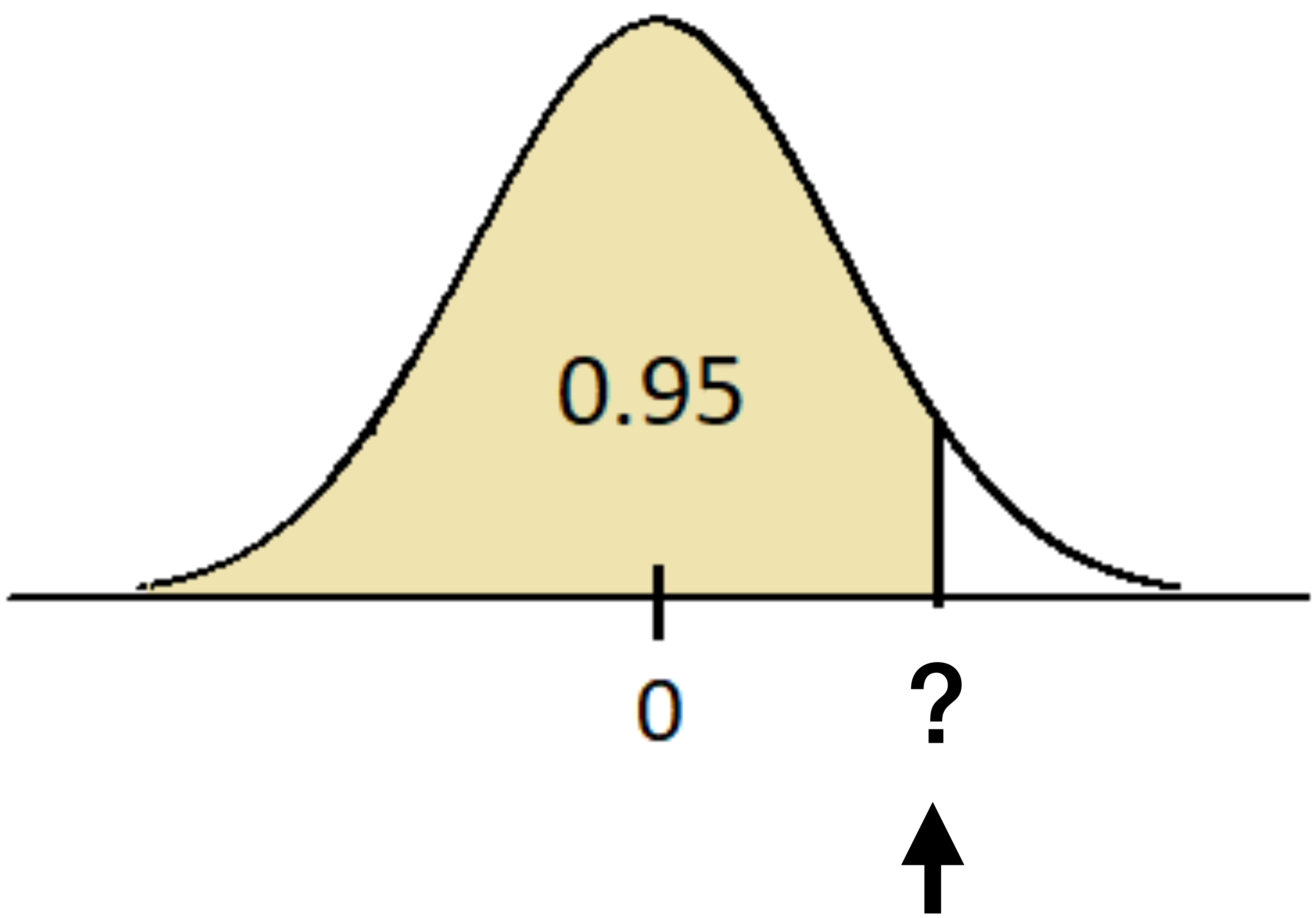


Example

Find a 90% confidence interval for the true average dollar amount that all customers in the region spent on organic produce each month.

We have

- $n = 200$
- $\bar{x} = 36$
- $\sigma^2 = 5$



qnorm(0.95) gives 1.645

Our Formula: $$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\alpha = 0.1 \qquad z_{\alpha/2} = z_{0.05} = 1.645$$

$$36 \pm 1.645\frac{\sqrt{5}}{\sqrt{200}}$$

$$\Downarrow$$

$$(35.74, 36.26)$$