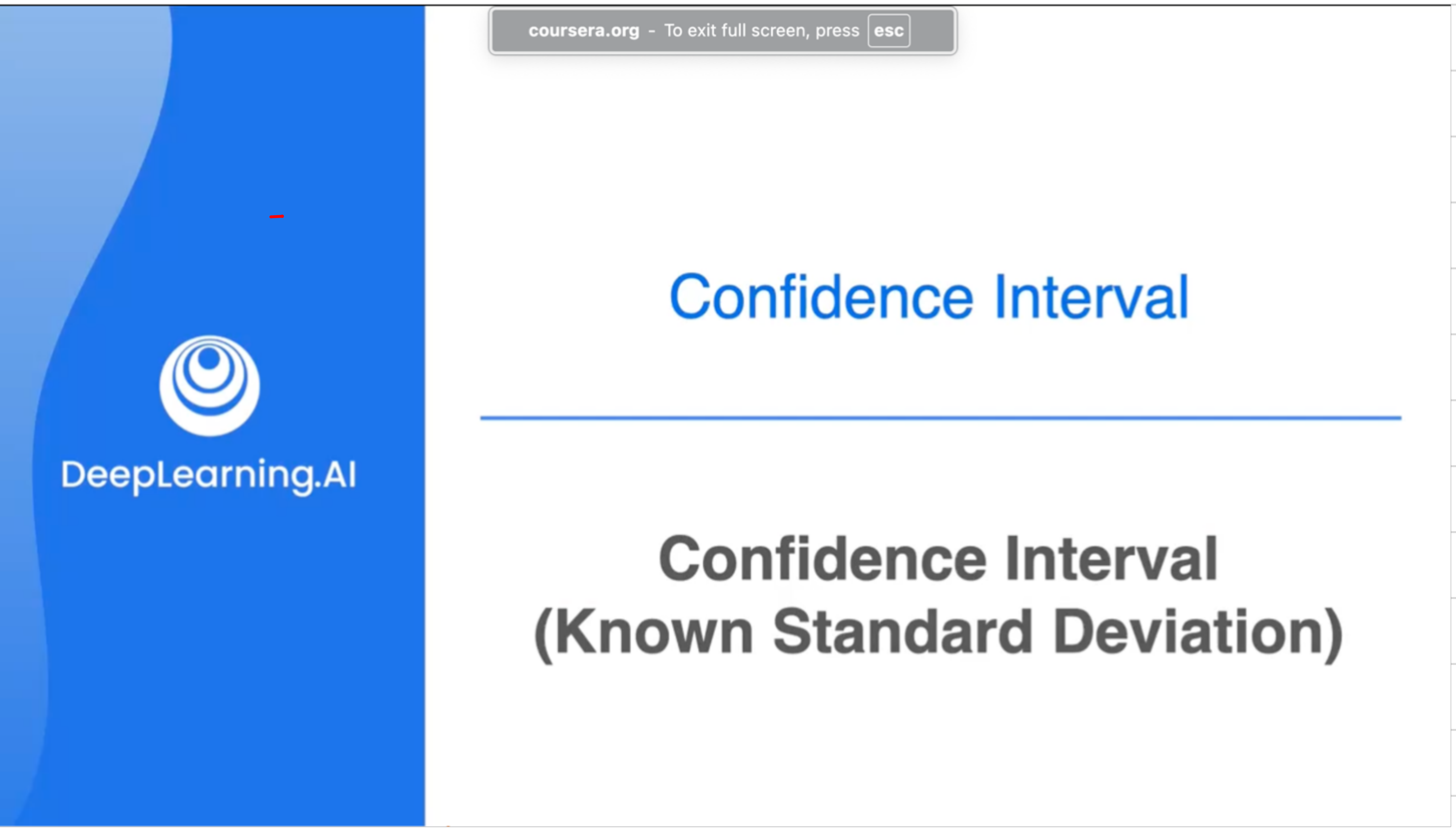
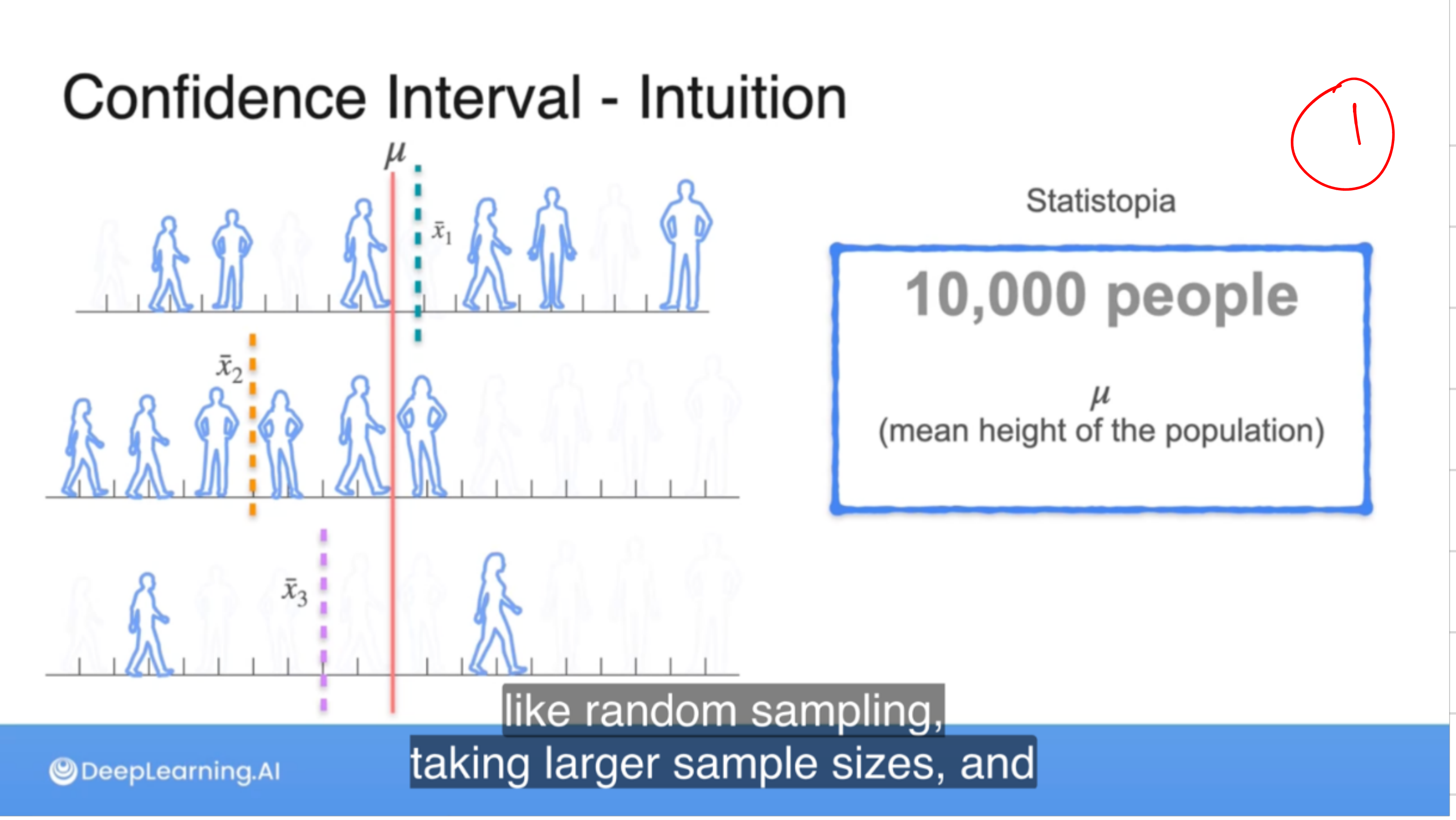
Day-65, Feb 3, 2025 (Magh 21, 2081 B.S.)

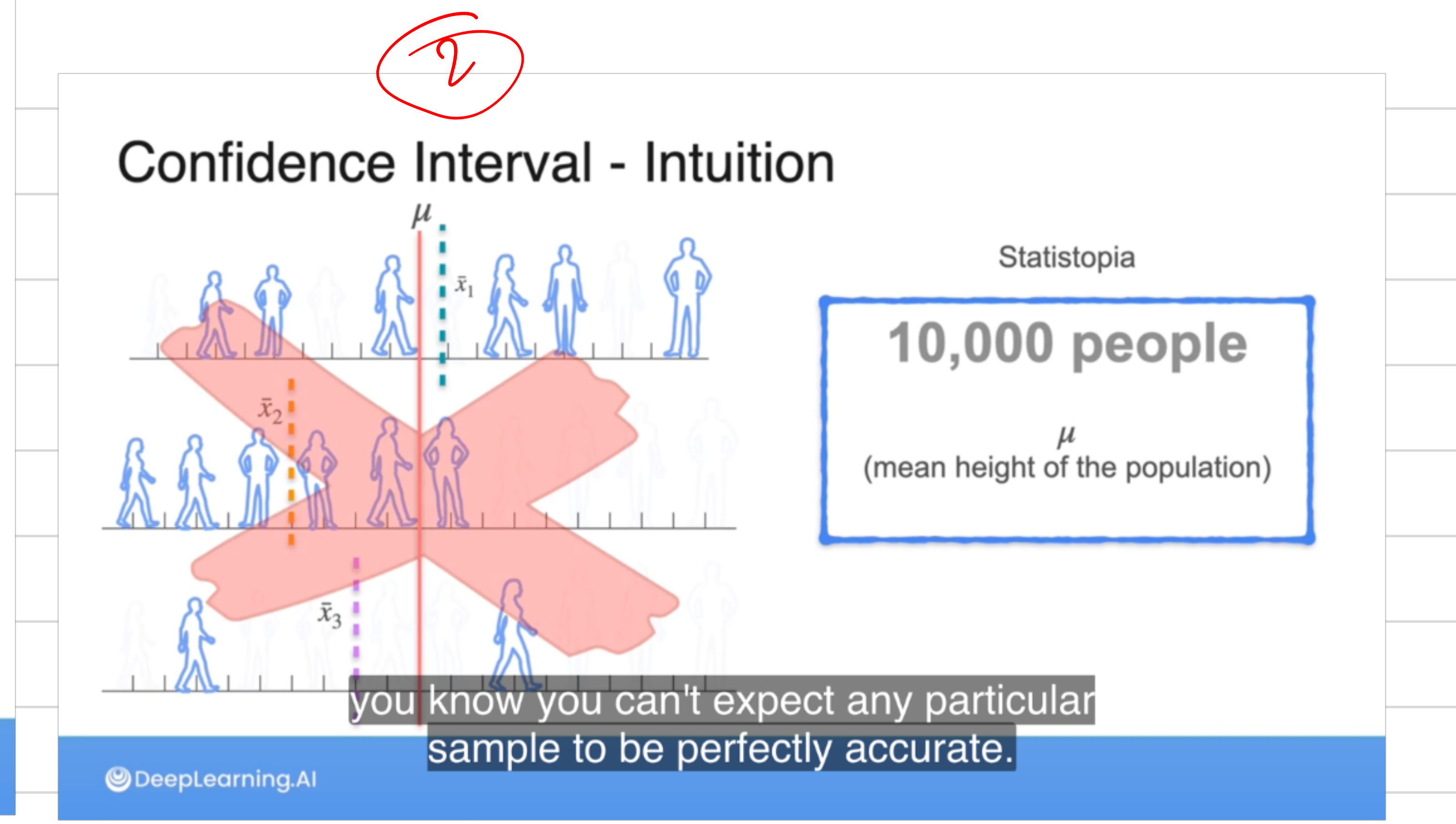
Confidence Interval.

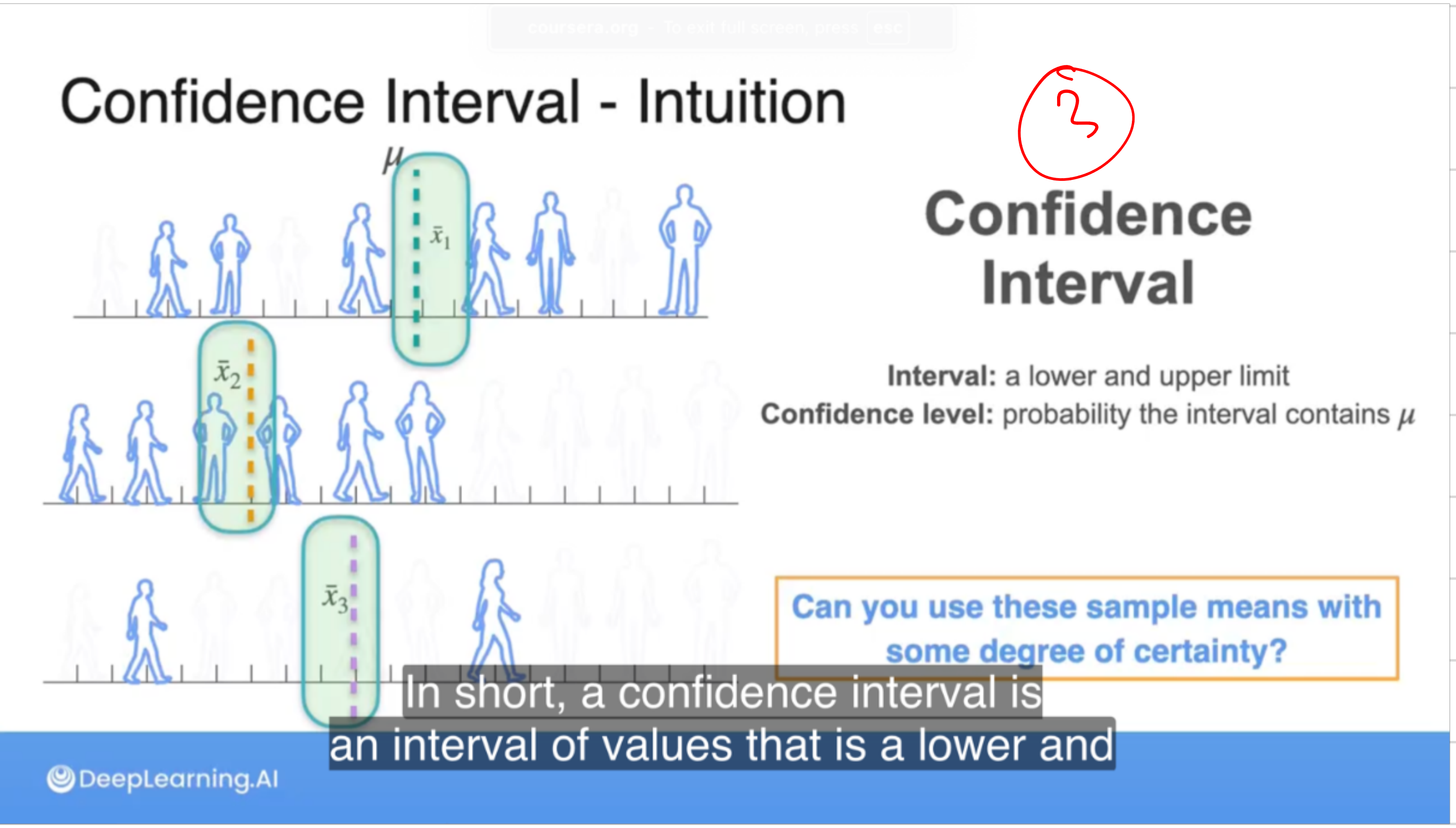
C.I.

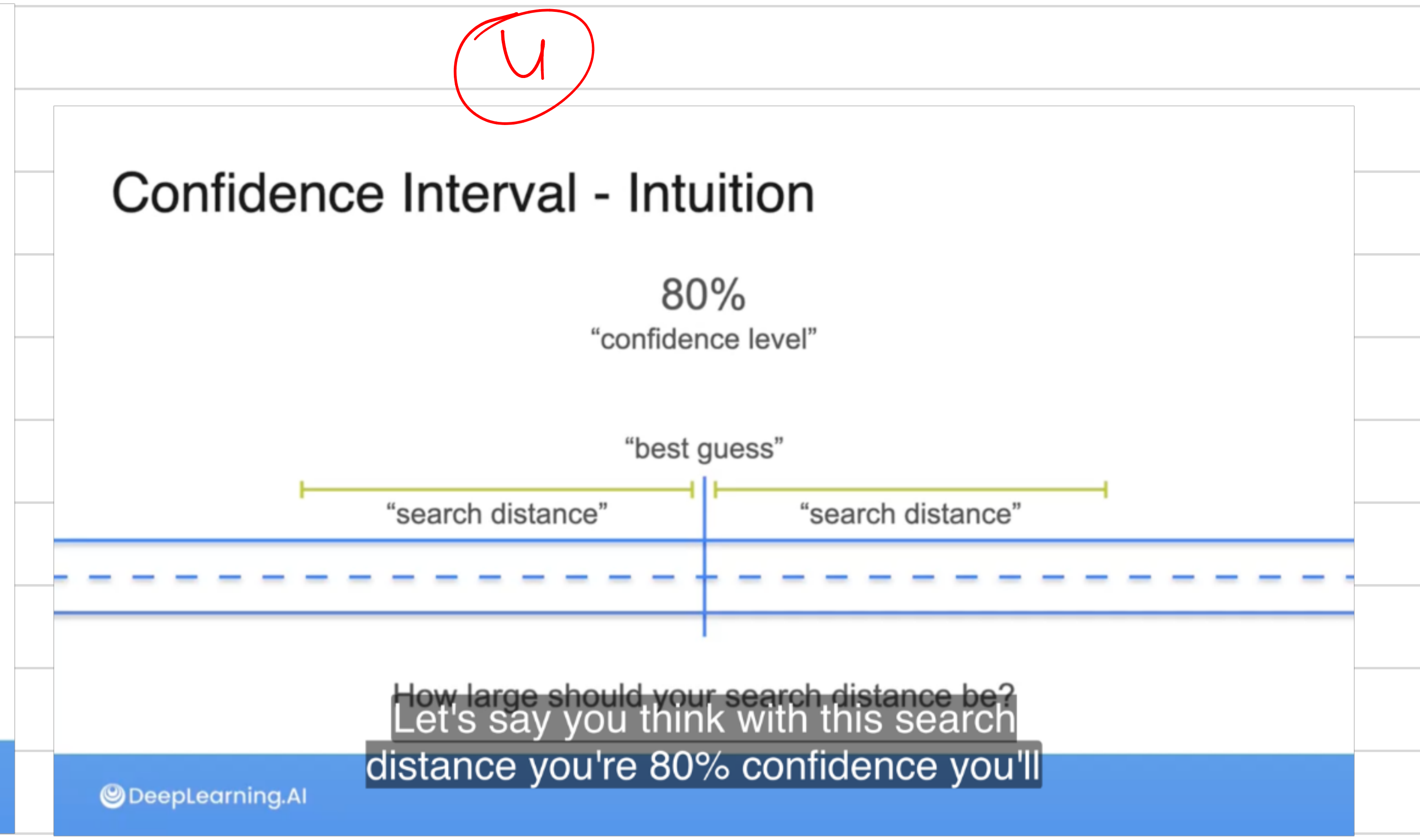
95%, 99%, 90%.

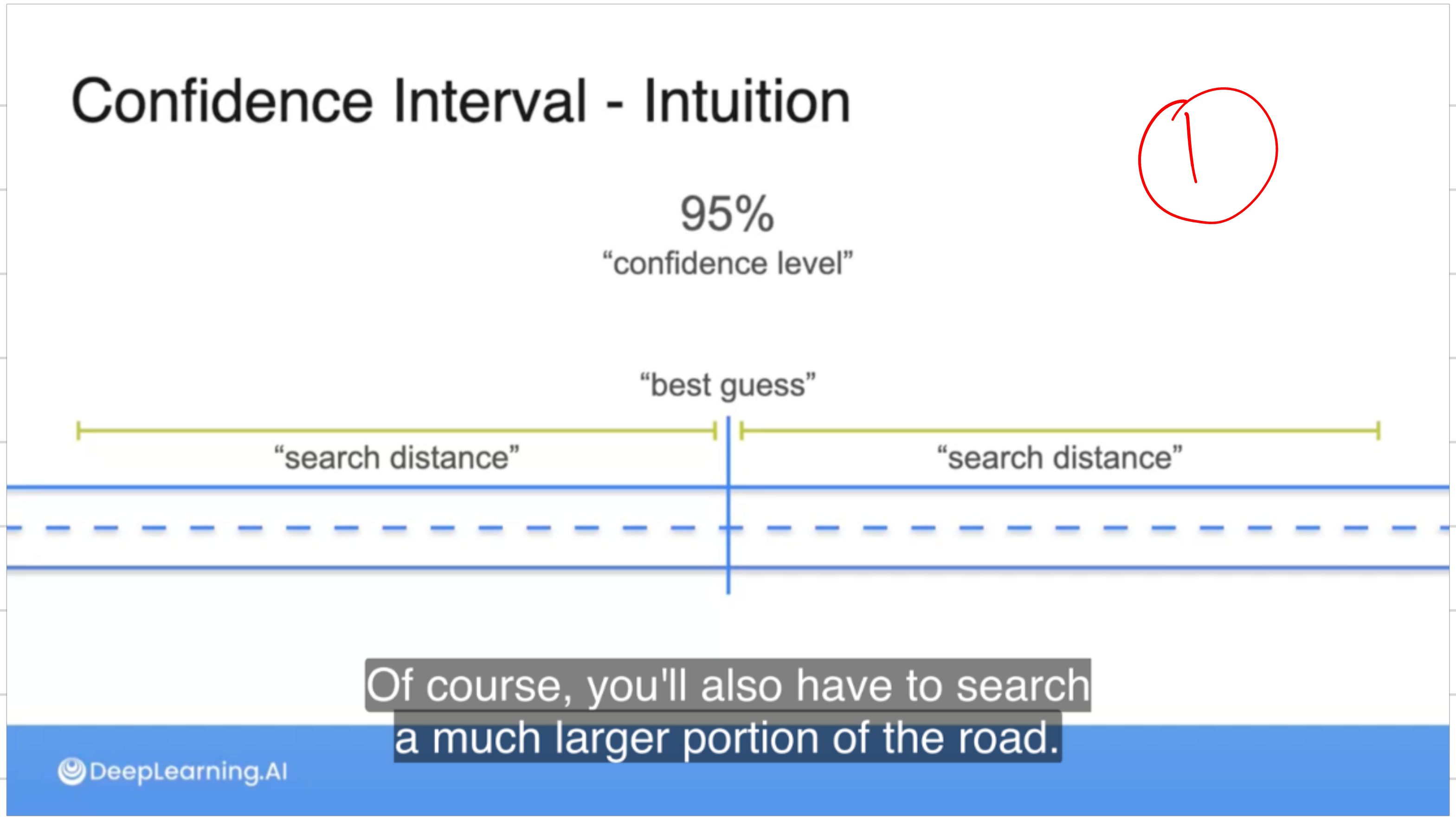


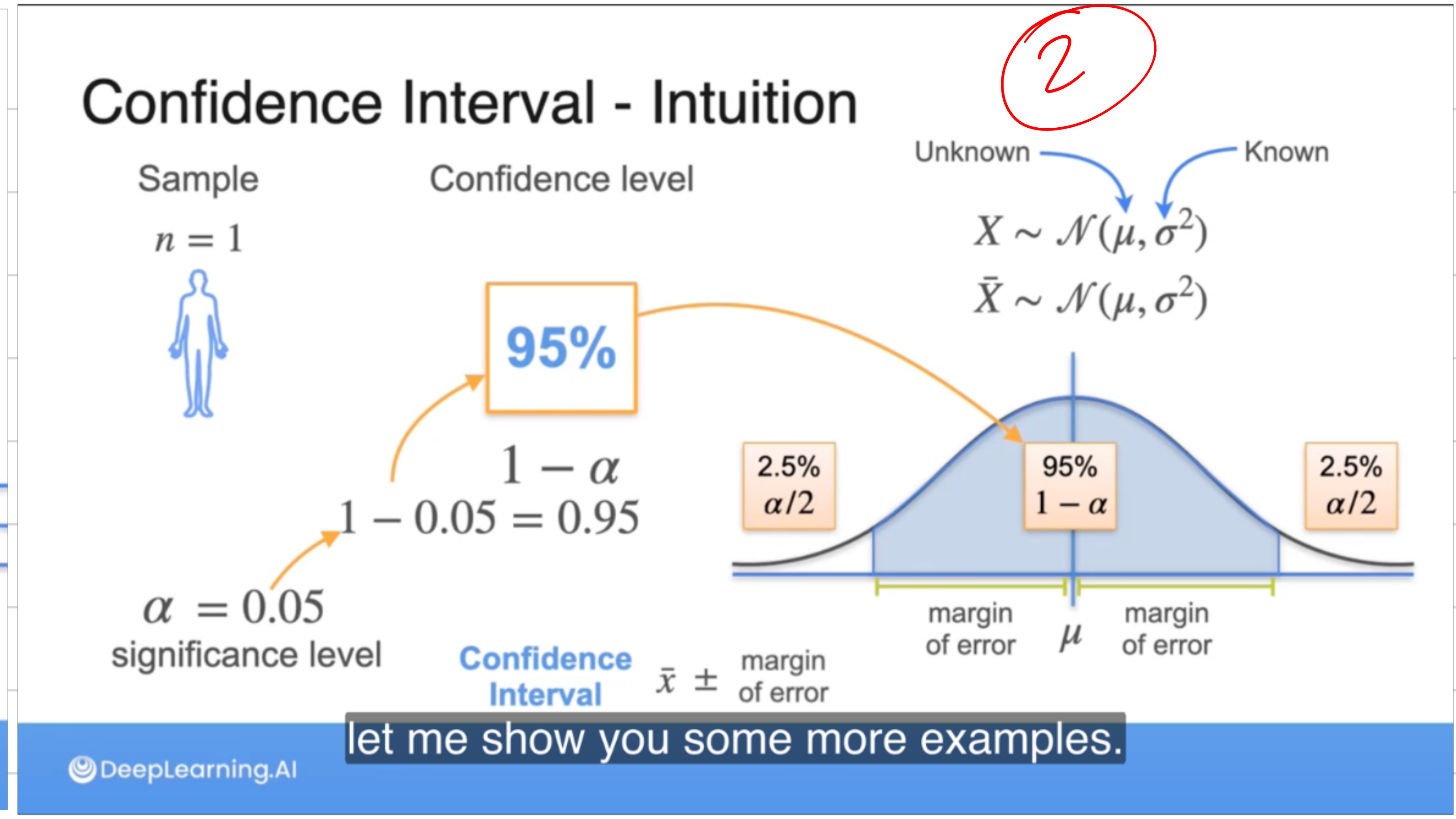

1
Confidence Interval - Intuition
μ
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
Statistopia
10,000 people
μ
(mean height of the population)
like random sampling,
taking larger sample sizes, and
DeepLearning.AI

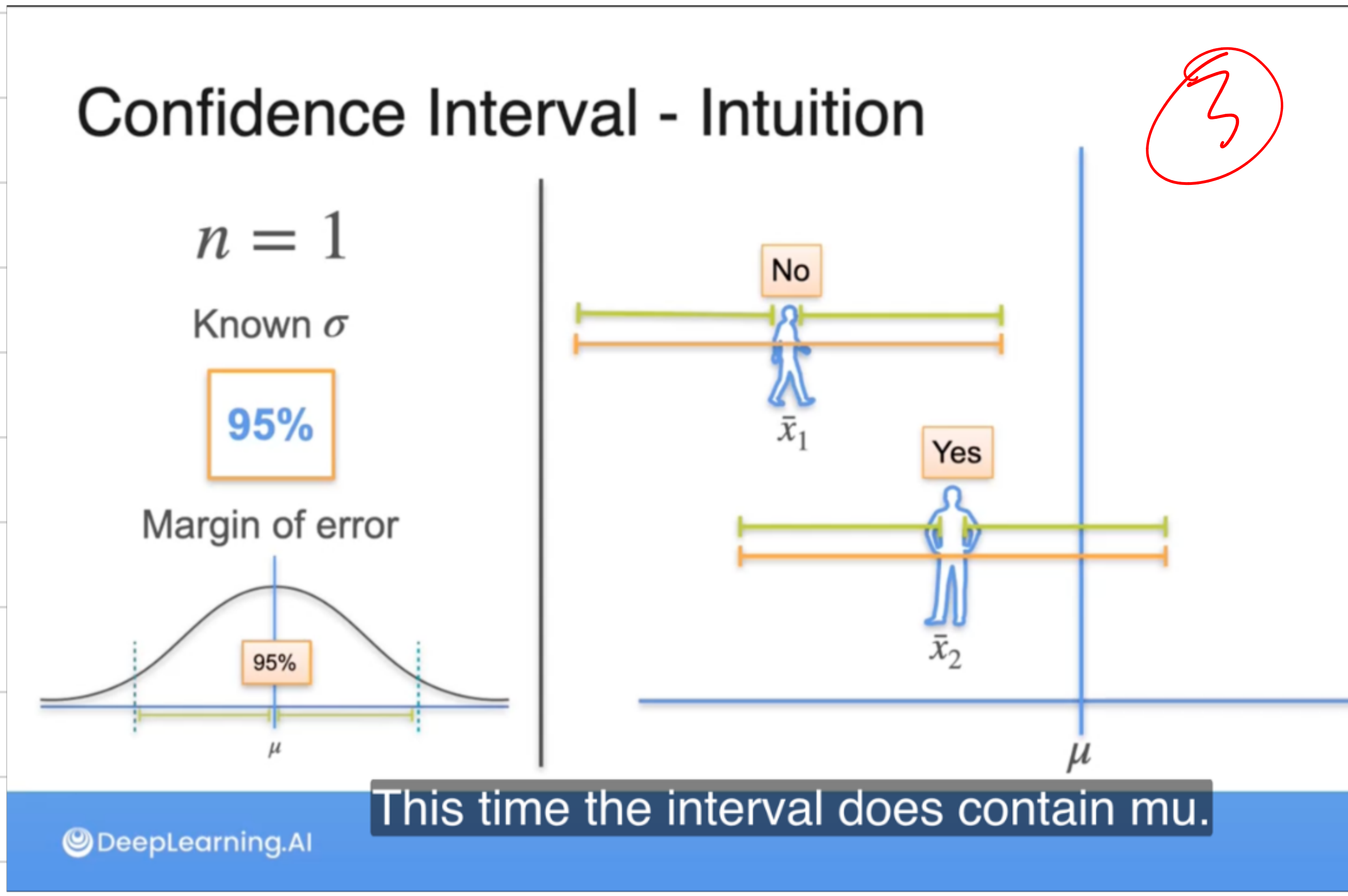

2
Confidence Interval - Intuition
μ
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
Statistopia
10,000 people
μ
(mean height of the population)
you know you can't expect any particular
sample to be perfectly accurate.
DeepLearning.AI

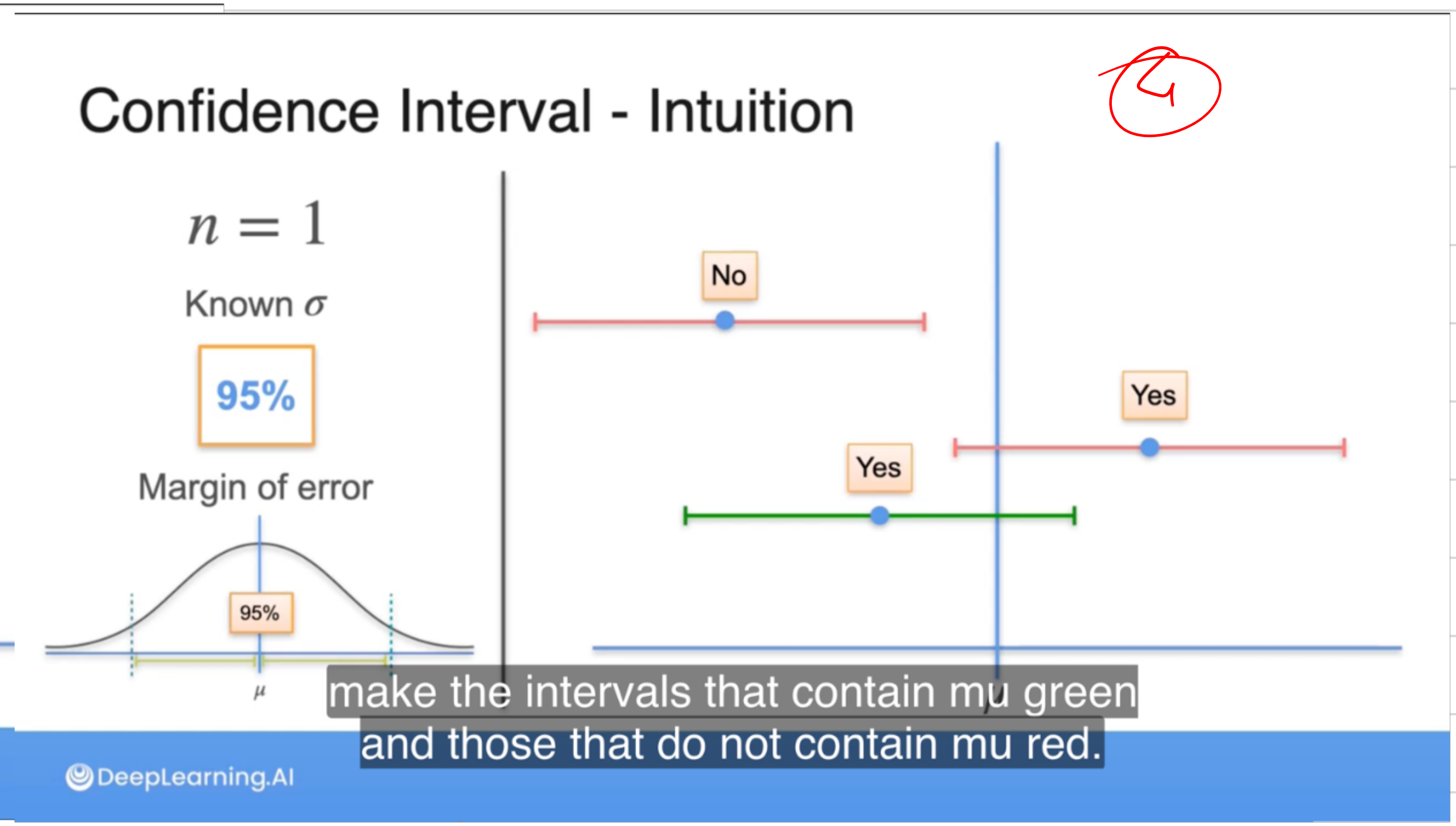

3
Confidence Interval - Intuition
μ
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
Confidence Interval
Interval: a lower and upper limit
Confidence level: probability the interval contains μ
Can you use these sample means with some degree of certainty?
In short, a confidence interval is
an interval of values that is a lower and
DeepLearning.AI

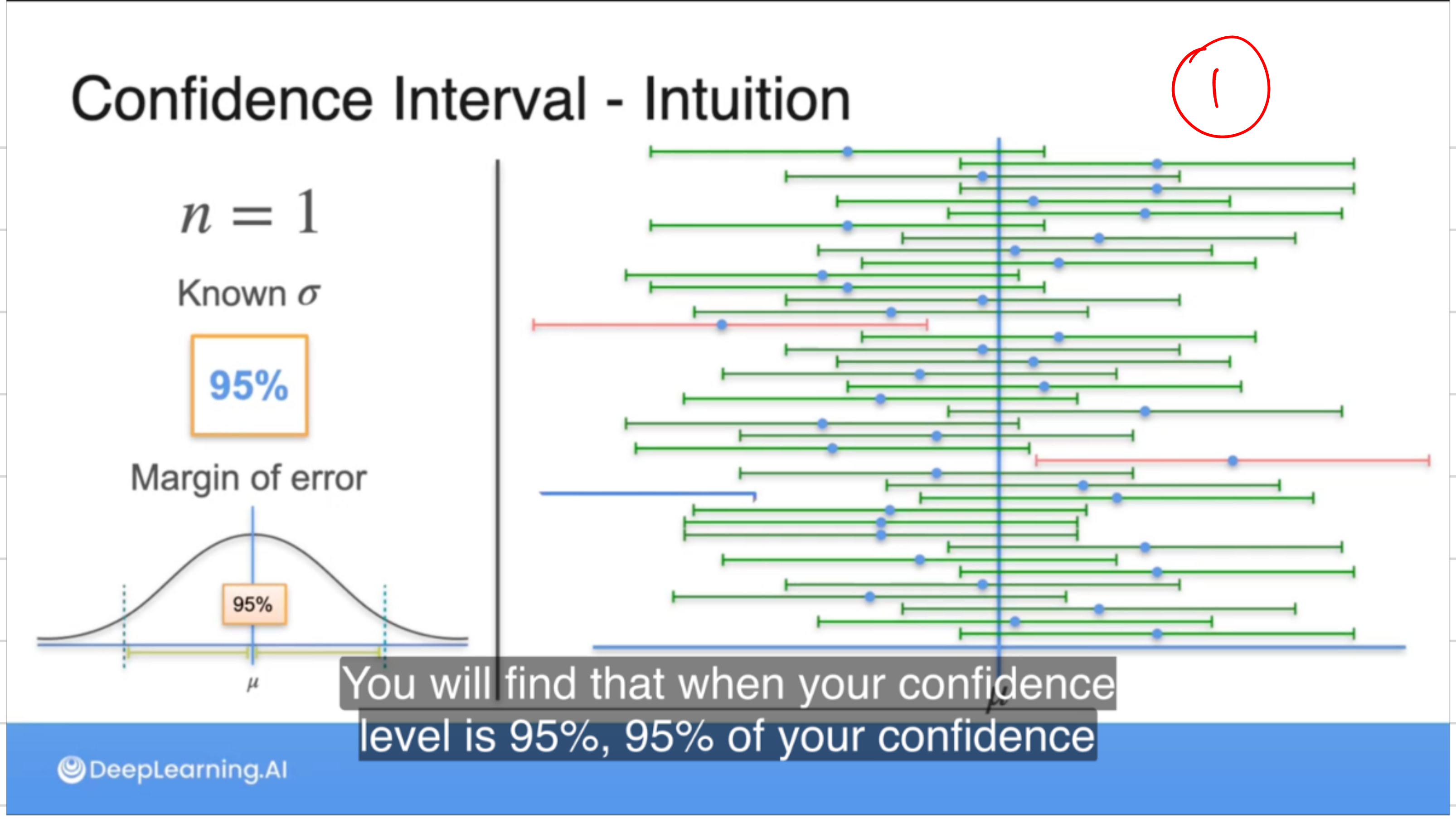

4
Confidence Interval - Intuition
80%
"confidence level"
"best guess"
"search distance"
"search distance"
How large should your search distance be?
Let's say you think with this search
distance you're 80% confidence you'll
DeepLearning.AI

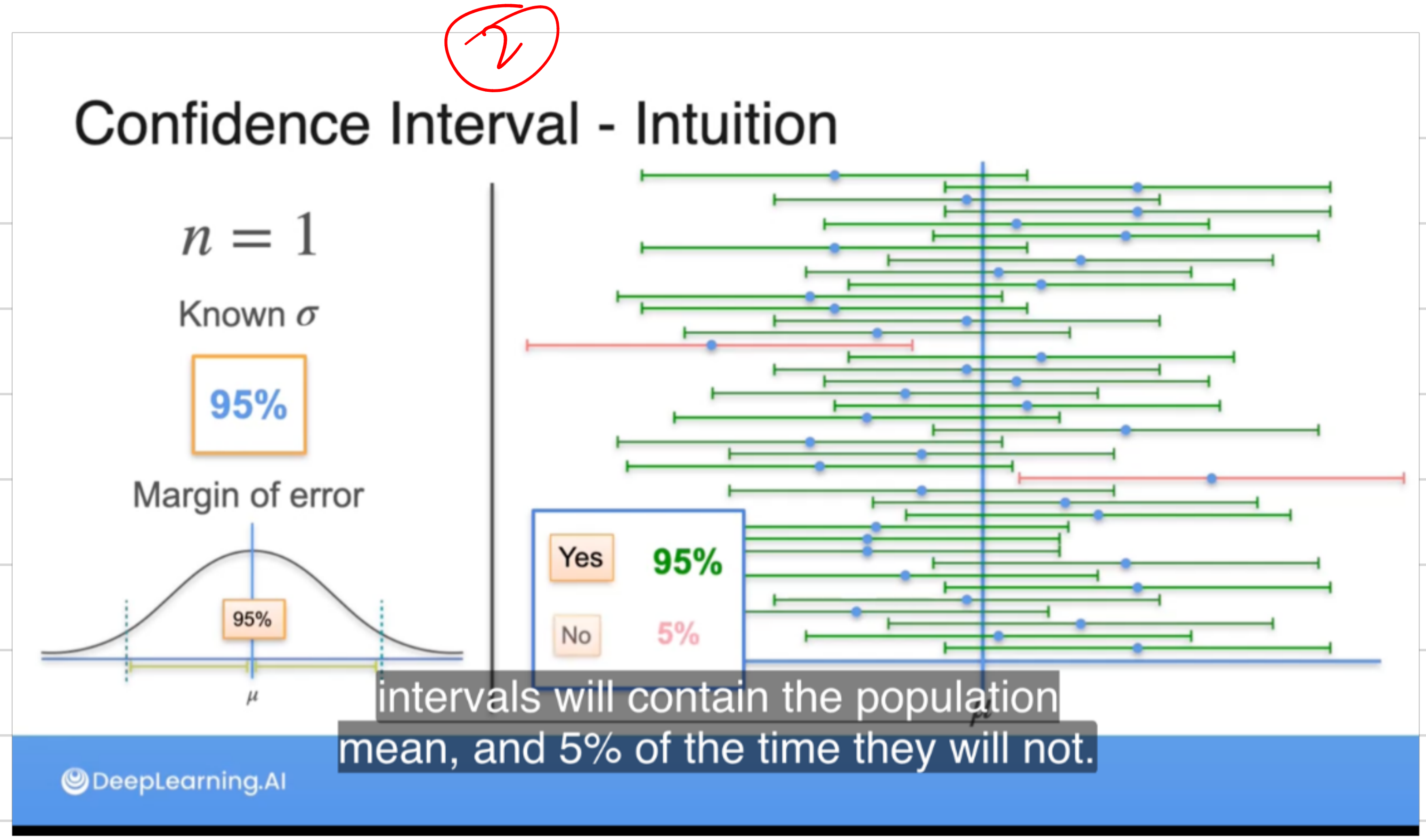

Confidence Interval - Intuition
1
95%
"confidence level"
"best guess"
"search distance"
"search distance"
Of course, you'll also have to search
a much larger portion of the road.
DeepLearning.AI

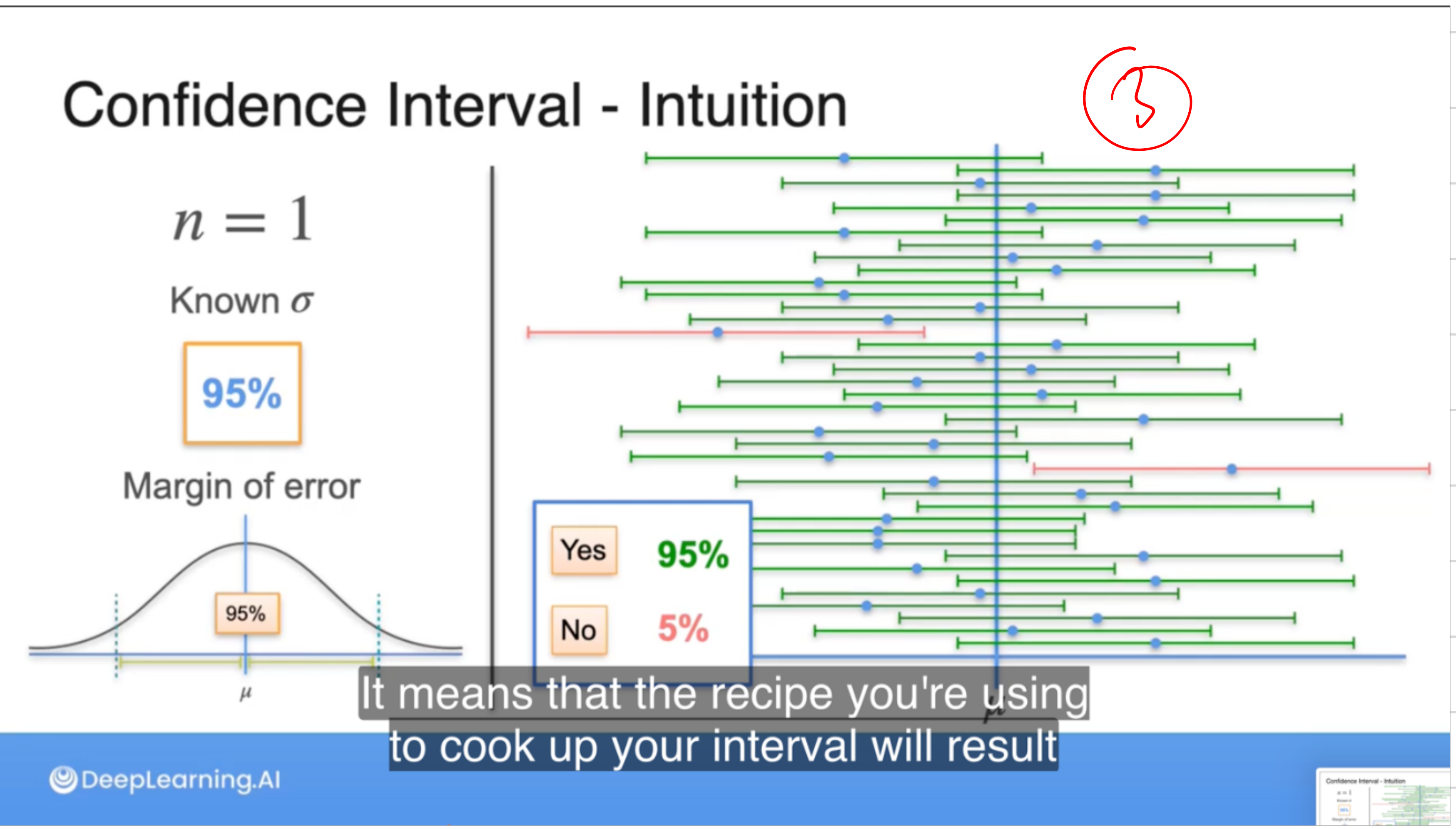

Confidence Interval - Intuition
2
Sample
$n = 1$
Confidence level
Unknown
Known
$X \sim \mathcal{N}(\mu, \sigma^2)$
$\bar{X} \sim \mathcal{N}(\mu, \sigma^2)$
95%
$1 - \alpha$
$1 - 0.05 = 0.95$
$\alpha = 0.05$
significance level
2.5%
$\alpha/2$
95%
$1 - \alpha$
2.5%
$\alpha/2$
margin of error
μ
margin of error
Confidence Interval
$\bar{x} \pm$ margin of error
let me show you some more examples.
DeepLearning.AI


Confidence Interval - Intuition
3
$n = 1$
Known σ
95%
Margin of error
95%
μ
No
$\bar{x}_1$
Yes
$\bar{x}_2$
μ
This time the interval does contain mu.
DeepLearning.AI


Confidence Interval - Intuition
4
$n = 1$
Known σ
95%
Margin of error
95%
μ
No
Yes
Yes
make the intervals that contain mu green
and those that do not contain mu red.
DeepLearning.AI







The Red ones are the ones that doesn't touch the μ or the mean but the 95% does. More / Large Sample Data gives us High CI.

Explanation uses real-world intuition (like searching for a lost key) and probability theory to explain the concept of confidence intervals in statistics. Here's a concise breakdown of the key ideas:

Uncertainty in Estimation:
When estimating a population mean (μ) from a sample, there's always some uncertainty since the true mean is unknown.
You rely on a sample mean ($\bar{x}$) to estimate μ.

Confidence Interval (CI):
A confidence interval provides a range of values (lower and upper limits) that are likely to contain μ with a specified confidence level (e.g., 95%).

The Key Analogy:
The road represents the range of possible values for μ.
Your best guess (where you park the car) represents $\bar{x}$.
The search distance (the interval width) reflects your confidence level: a wider search distance increases confidence but requires more effort.

Confidence Level:
A 95% confidence level means that 95% of confidence intervals generated in repeated sampling will contain μ.
There's still a 5% chance the interval won't contain μ, split equally between the two tails of the distribution (2.5% each).

Formula for CI:

$\bar{x}$ ± margin of error

The margin of error depends on the standard deviation (σ) and the confidence level (linked to α, the significance level).

Visualizing Confidence Intervals:

When repeatedly sampling from the population:

Green intervals: Contain μ.

Red intervals: Do not contain μ.

For a 95% confidence level, 95 out of 100 intervals are expected to contain μ.

Understanding Significance Level (α) and Confidence Level:

Confidence Level = 1 – α.

For α =0.05, the confidence level is 95%.

In a normal distribution, $\alpha/2$

α/2 is on each tail (2.5% each).

Random Nature of Confidence Intervals:

You generate one interval from one sample.

In repeated sampling, 95% of intervals will contain the true mean, and 5% will not.

Visualizing red and green intervals reinforces the fact that not every interval captures μ.

CI (Range)

key statistical concepts related to confidence intervals, sample size, and their impact on precision. Below is a summarized breakdown of the main points:

Key Concepts:
Confidence Interval Ingredients:

A sample mean ($\bar{x}$) and a margin of error around it.
The confidence level, typically 95%, indicates the probability that the interval contains the true population mean, μ.
Effect of Sample Size (n):

Increasing sample size (n) reduces the standard deviation of the sample mean ($\sigma/\sqrt{n}$).
This leads to a narrower confidence interval while maintaining the same confidence level, making estimates more precise.
Impact of Confidence Level:

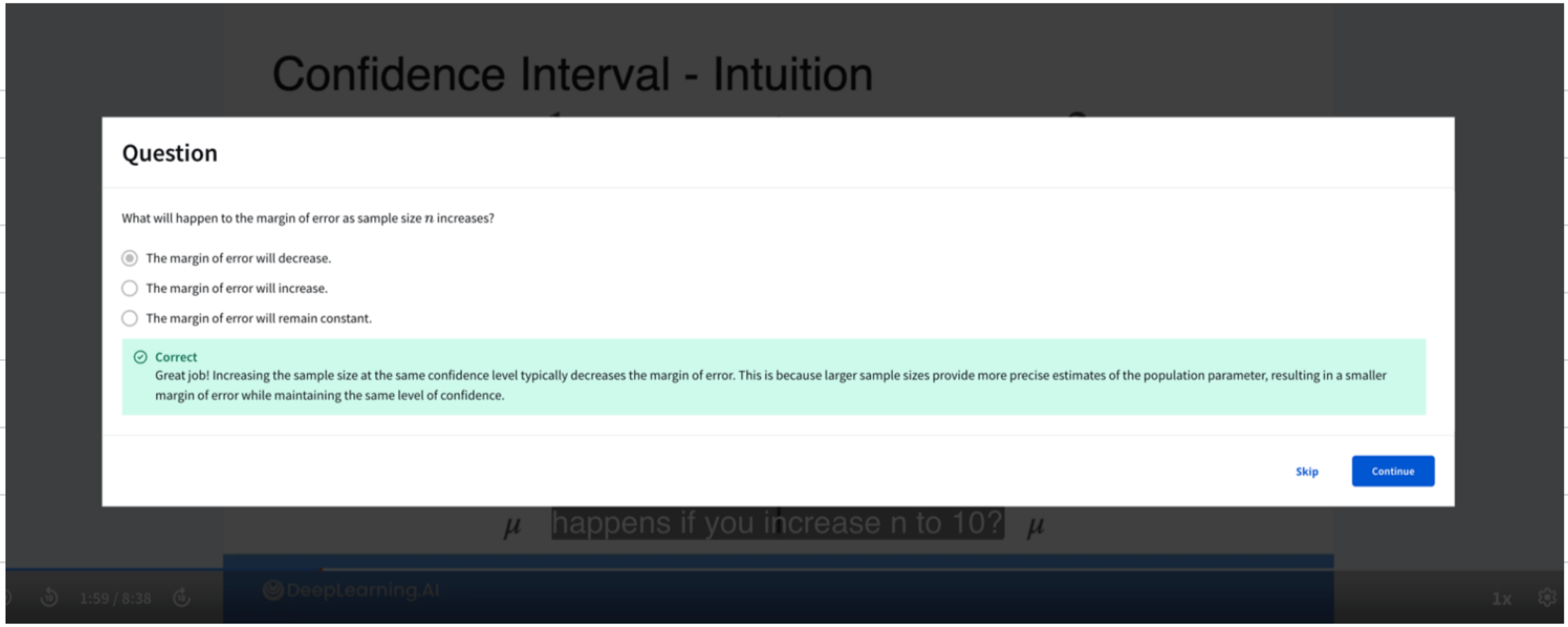
Higher Confidence Level (95% or 99%) requires larger margins of error to capture more of the sampling distribution.
Lower Confidence Level (e.g., 70%) results in smaller margins of error, but intervals are less likely to contain the population mean.
Balancing Trade-offs:

Smaller margins of error with a high confidence level aren't free — they require larger sample sizes.
A practical approach is to aim for a confidence level of 95%, which balances precision and reliability.
Visual Insights:

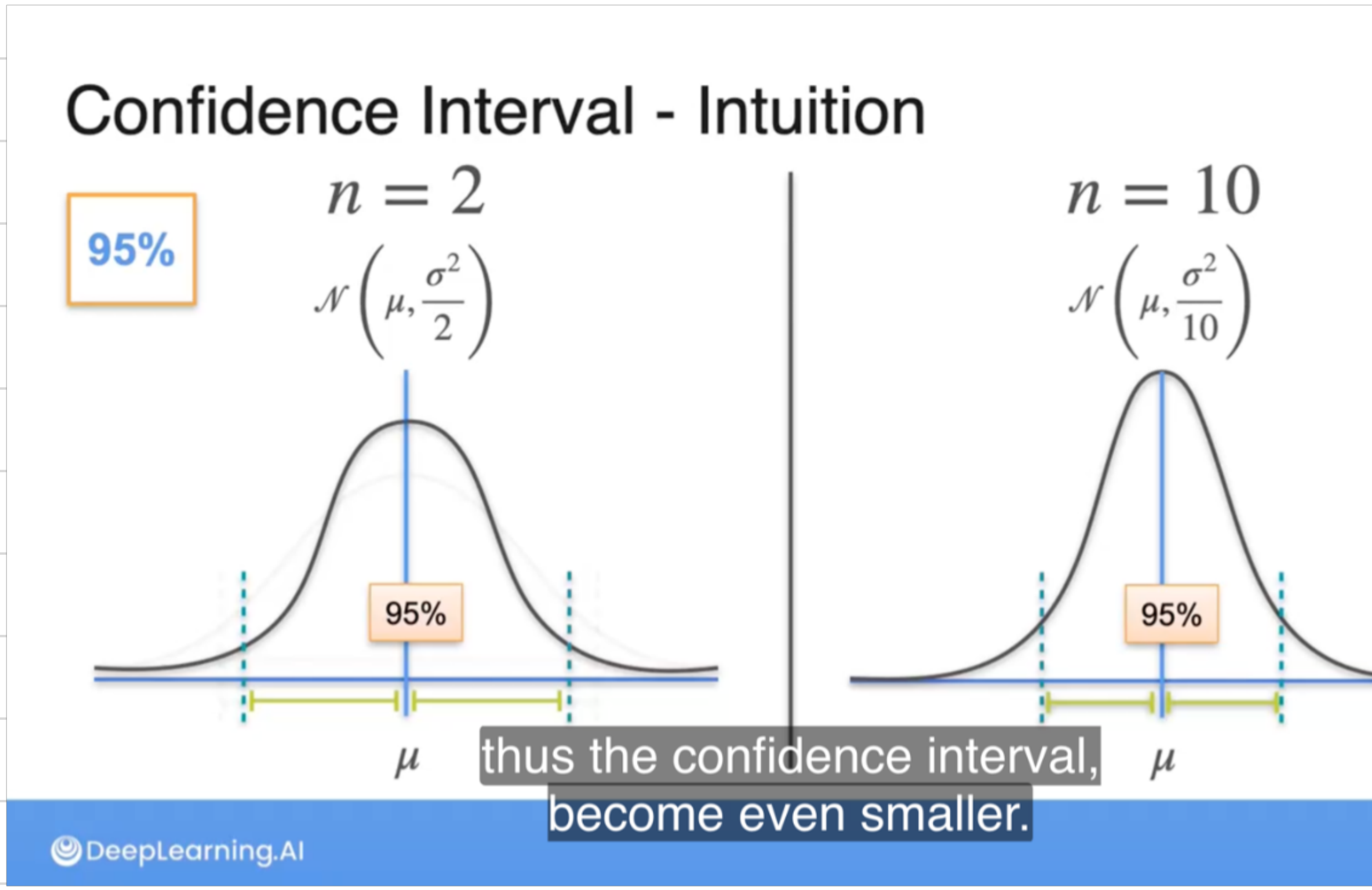
Simulations demonstrate that as n increases, confidence intervals shrink.
Lower confidence levels have shorter intervals but a higher risk of not containing μ.
Real-World Application:

When estimating population parameters, it's ideal to have high confidence and narrow intervals, achievable by gathering more data.

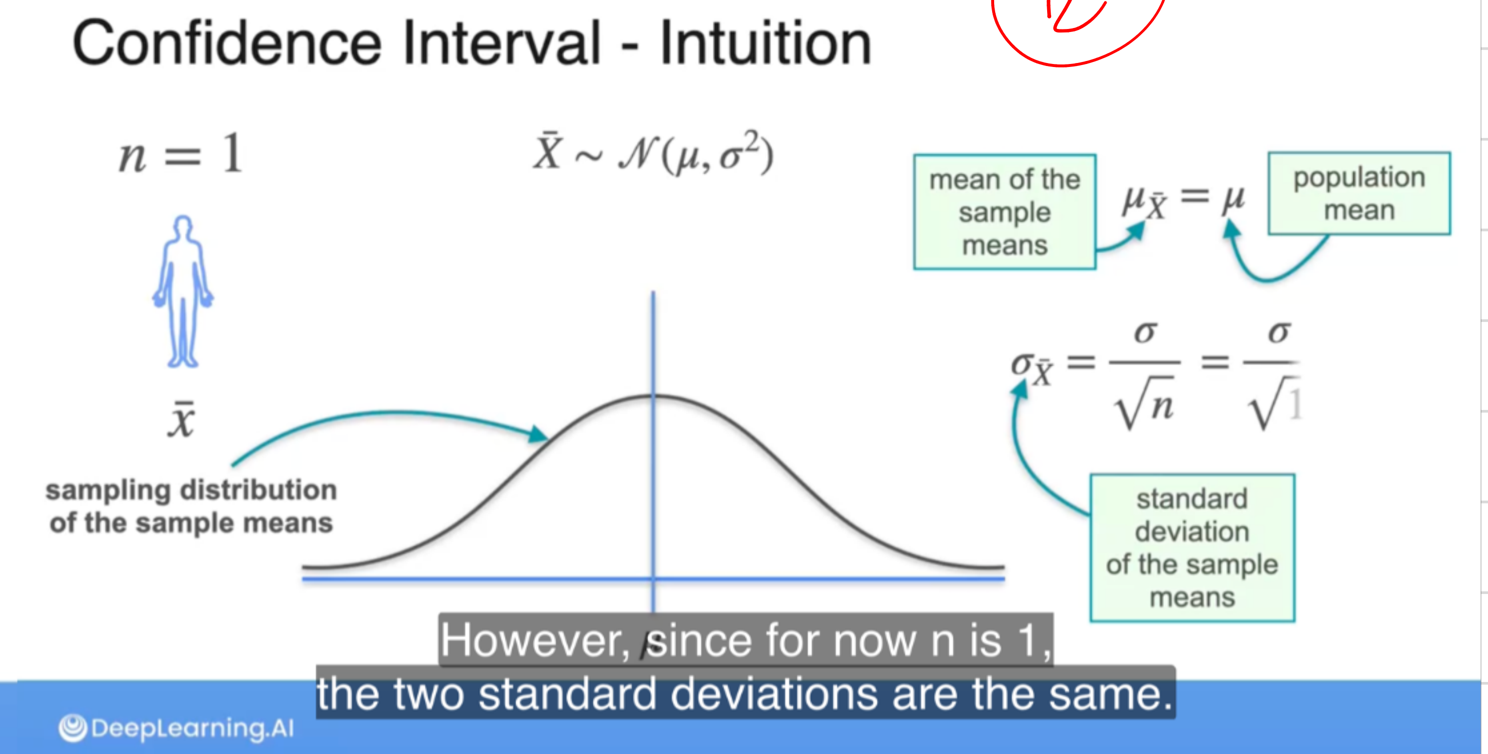
Margin of Error $\Downarrow$ if $\Uparrow$ Sample Size Increases!

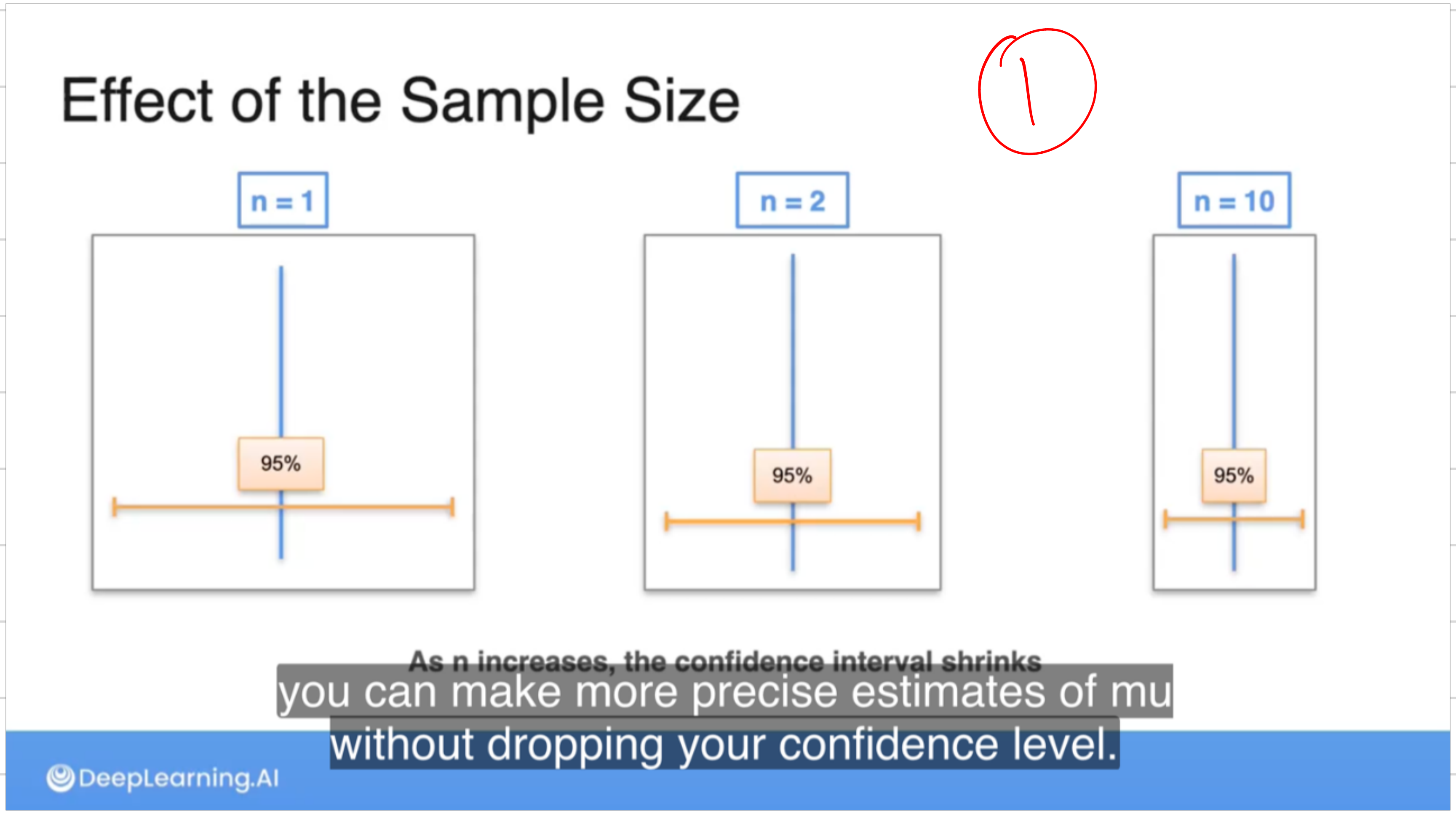


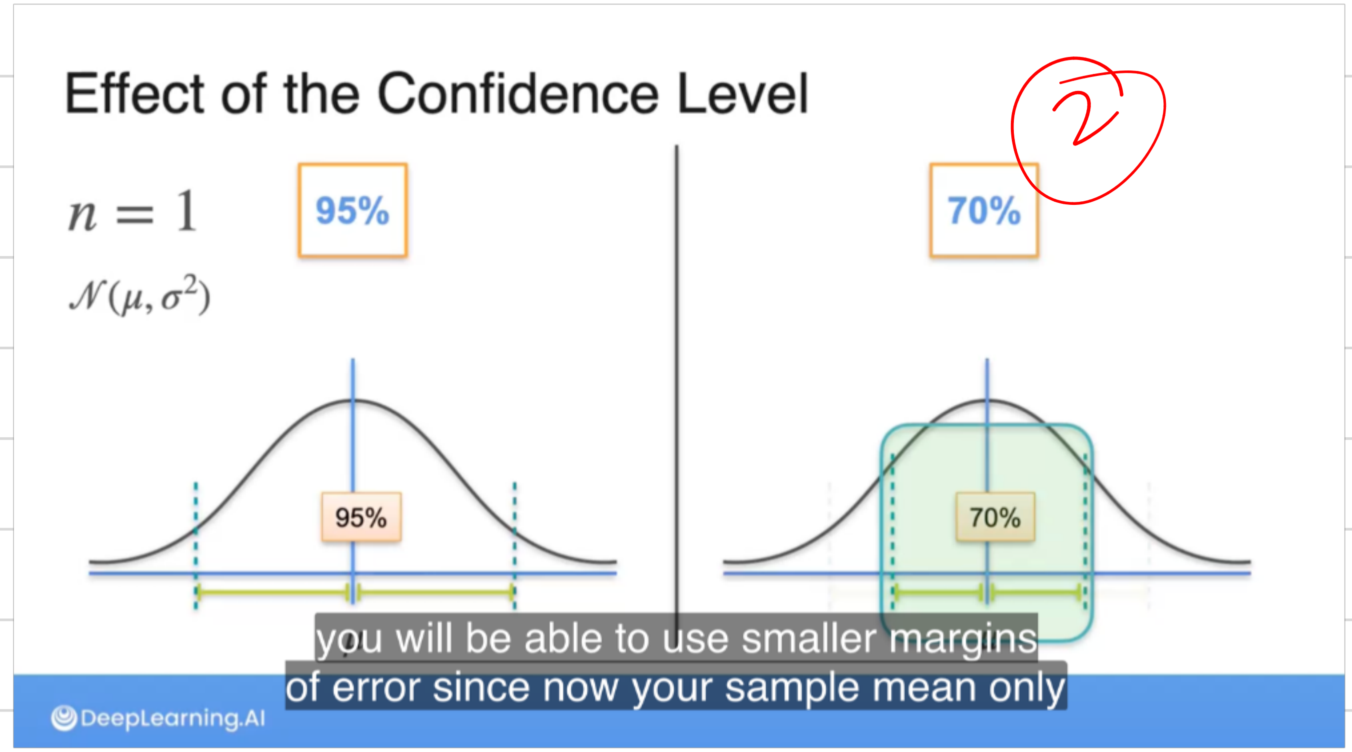
1

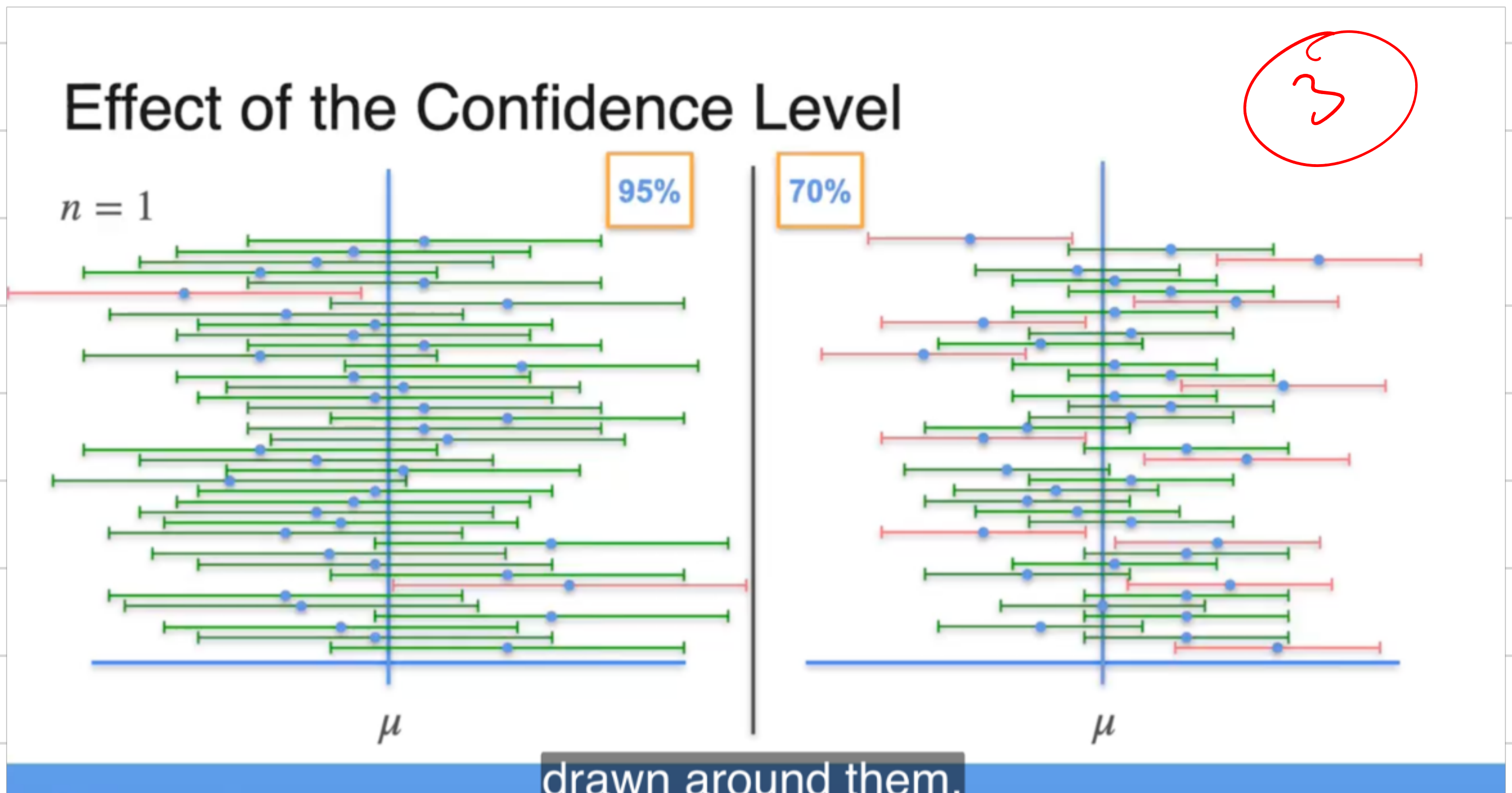


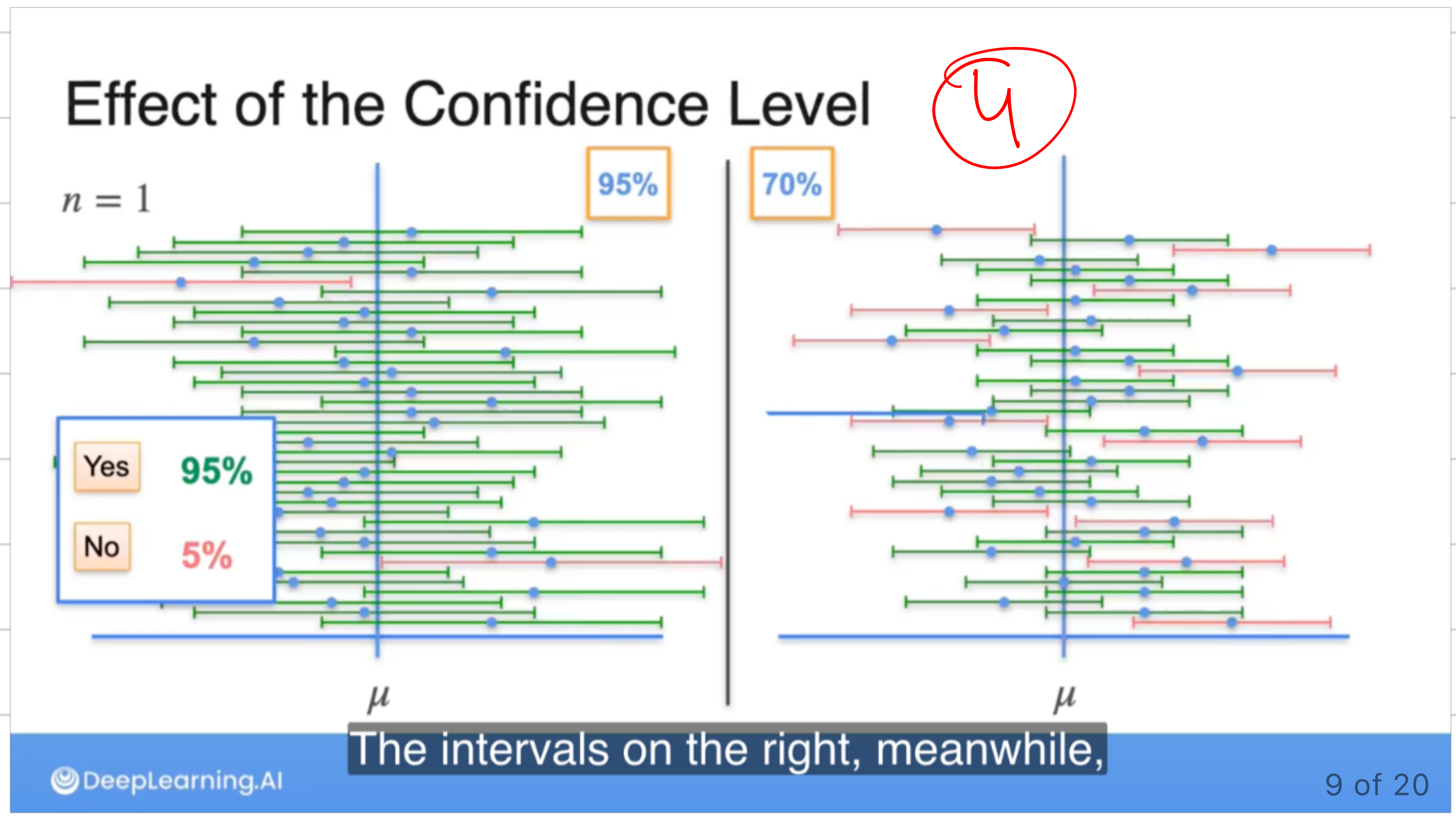
2

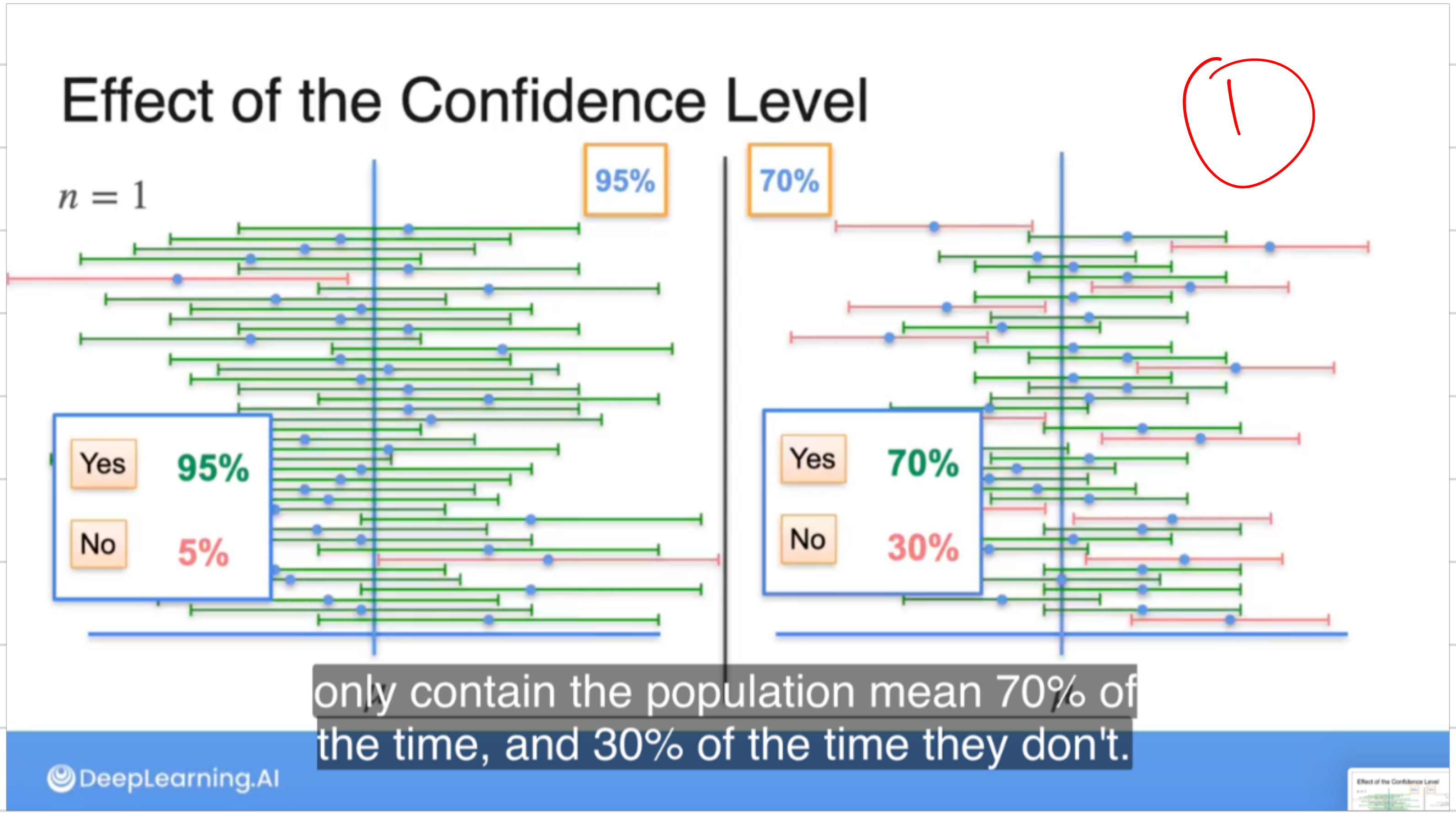


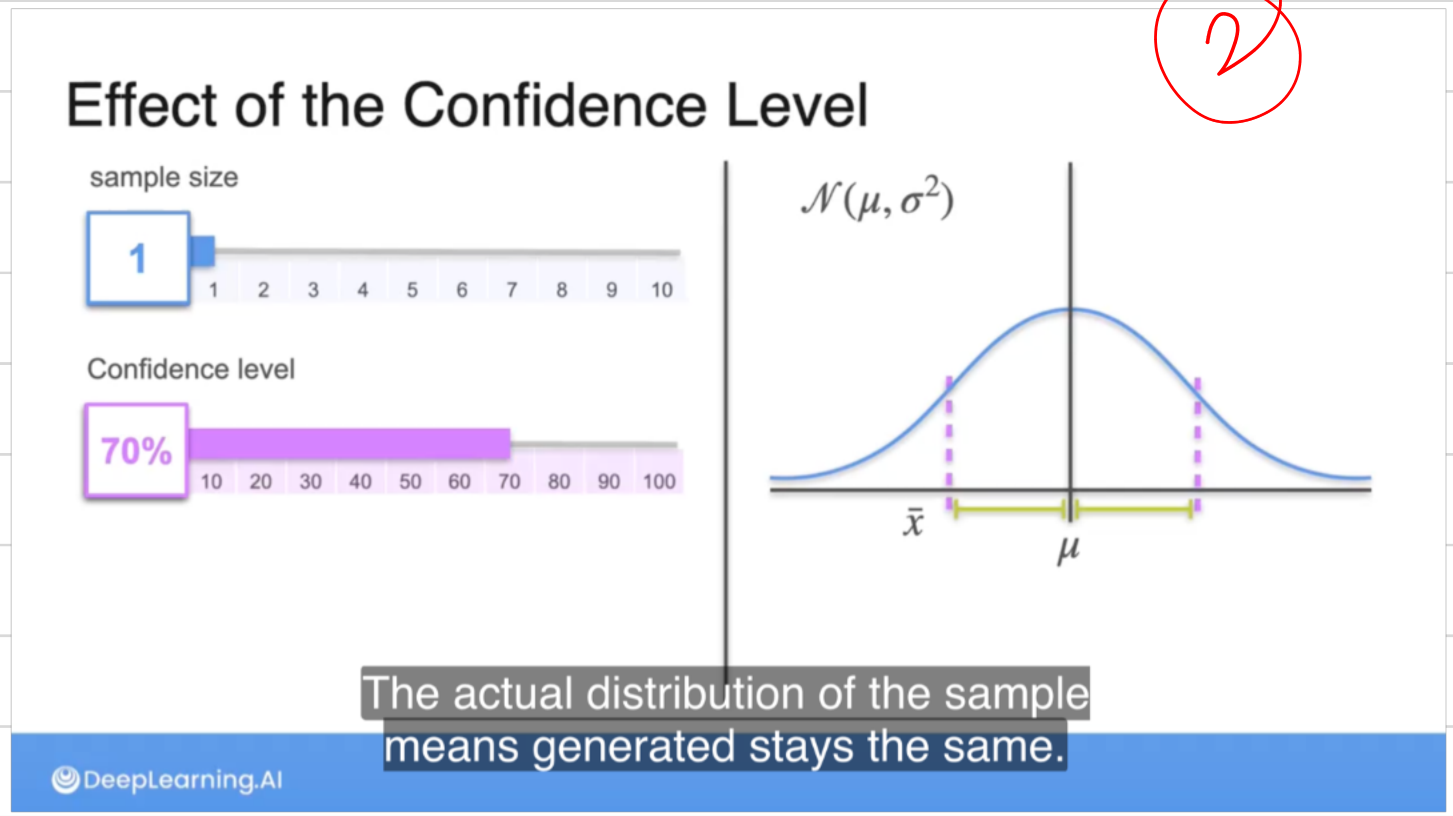

Effect of the Sample Size
1
n = 1
n = 2
n = 10
95%
95%
95%
As n increases, the confidence interval shrinks
you can make more precise estimates of mu
without dropping your confidence level.
DeepLearning.AI

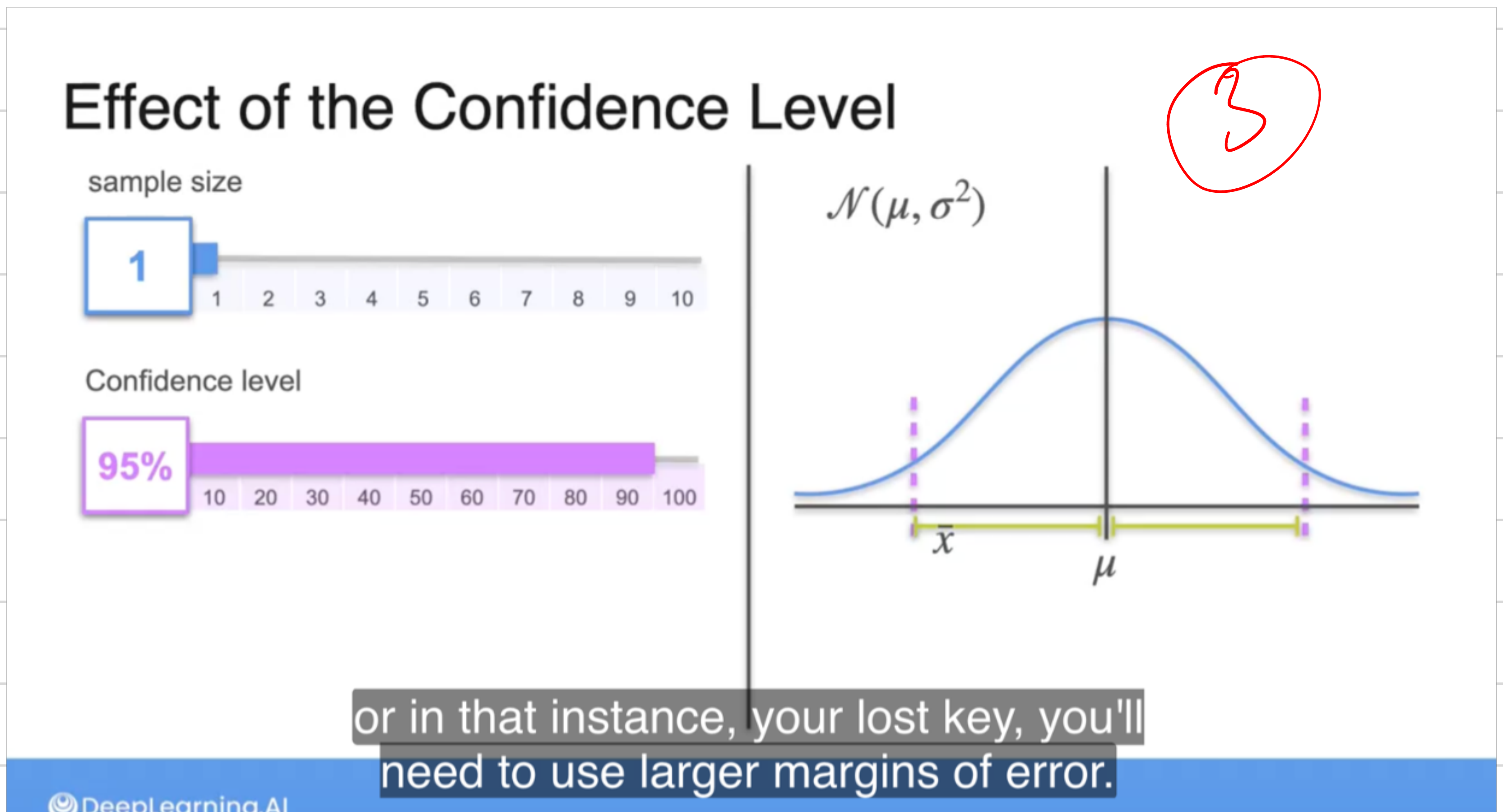

Effect of the Confidence Level
2
$n = 1$
95%
70%
$\mathcal{N}(\mu, \sigma^2)$
95%
70%
you will be able to use smaller margins
of error since now your sample mean only
DeepLearning.AI


Effect of the Confidence Level
3
$n = 1$
95%
70%
μ
μ
drawn around them.


Effect of the Confidence Level
4
$n = 1$
95%
70%
Yes
95%
No
5%
μ
μ
The intervals on the right, meanwhile,
DeepLearning.AI







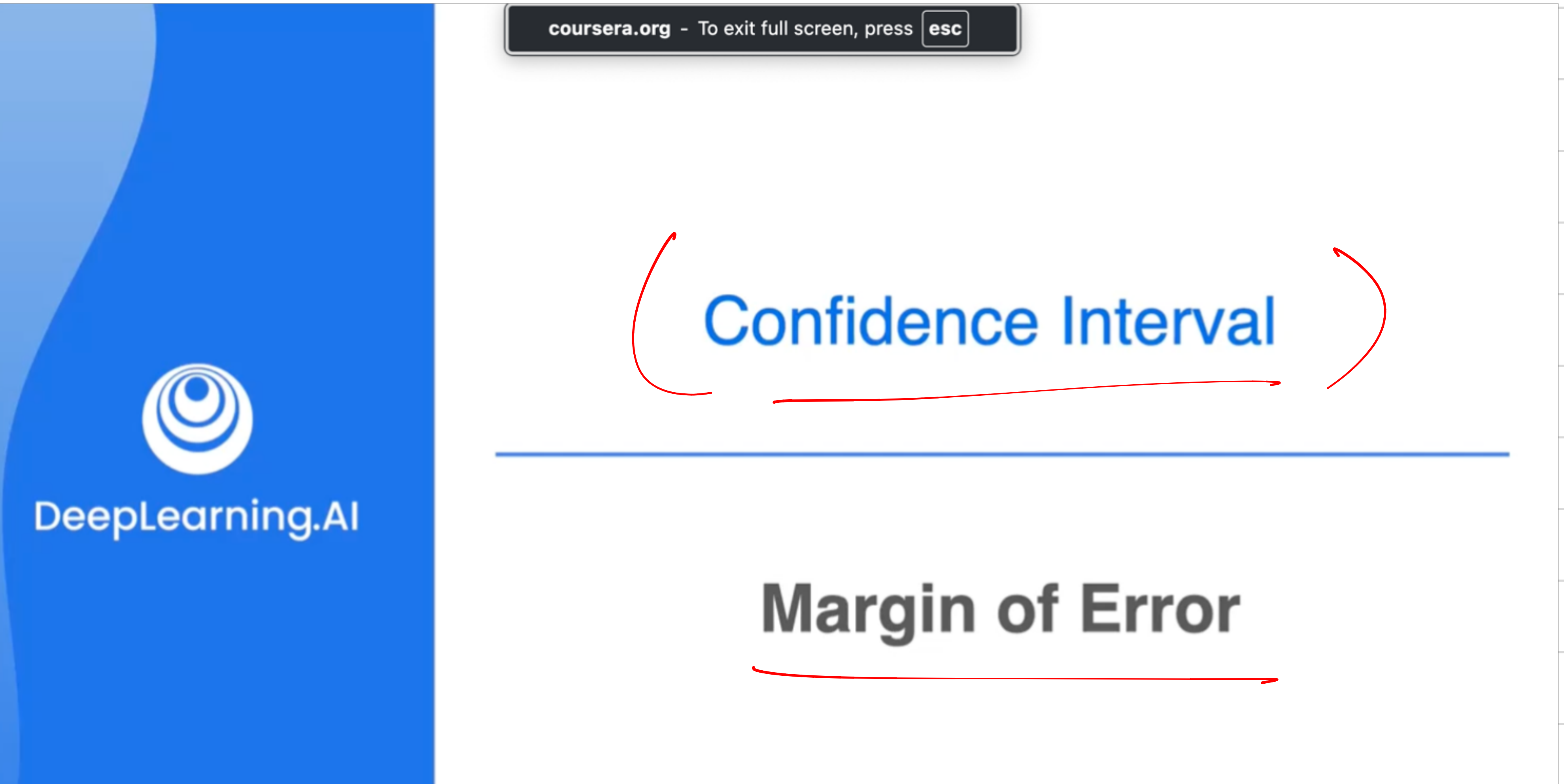
4

Summary

- Confidence intervals are a sample mean with a margin of error added to each side
- Confidence level: the probability a confidence interval contains μ (for example 95%)
- Ideally you have both high confidence and a narrow interval
- Larger samples (more data) will give a narrower interval
- Decreasing confidence level will also shrink the interval

So far, you've learned about all
these concepts at a high level.

DeepLearning.AI



Why Use Fixed Z-Scores?
To save time and maintain accuracy, statisticians have defined fixed Z-scores for the most common confidence levels:

90% Confidence Level: Z = 1.645
95% Confidence Level: Z = 1.96
99% Confidence Level: Z = 2.575
These fixed values simplify calculations without the need for recomputation every time.

Imagine you have a line, and on that line, there's a middle point where most of your toys are. Some toys are close to the middle, and some toys are far away.
Now, let's say you want to know how far the toys are from the middle. The Z-score is like a way to measure how far a toy is from the middle.

Here's an example:
If you say "I want to know how far a toy is from the middle and I want to include almost all my toys, 95% of them," the Z-score is like saying "go about 2 steps away from the middle to make sure you have most of the toys."
That 2 steps is what we call Z=1.96. It's just a special number that tells us how far we need to go from the middle to see most of the toys.

Why is it fixed?
We already know that 95% of the toys are within 2 steps of the middle. So, we use that number without needing to count all the toys again every time!

1

Key Steps in Confidence Interval Construction:

1. **Identify Key Ingredients:**
   - **Sample Mean ($\bar{x}$):** The average value from your sample.
   - **Margin of Error:** Determines how much variability exists around the sample mean.
2. **Calculate Margin of Error:**
   - Margin of Error $= Z \times \text{Standard Error}$
   - Standard Error $= \frac{\sigma}{\sqrt{n}}$
   - Z is the z-score corresponding to the desired confidence level (for 95%, $Z = 1.96$).
3. **Create Confidence Interval:**
   - Lower Limit $= \bar{x} - \text{Margin of Error}$
   - Upper Limit $= \bar{x} + \text{Margin of Error}$

Important Concepts:

- **Z-Scores and Critical Values:**
  - For a 95% confidence level, the critical values are -1.96 and 1.96.
  - Critical values exclude 5% of the distribution, 2.5% in each tail.
- **Normal Distribution Assumption:**
  - When the population follows a normal distribution, the sample mean distribution is also normal.
  - If the sample size n is large, the Central Limit Theorem ensures that the sample mean follows a normal distribution regardless of the population shape.
- **Generalization:**
  - Even if the population distribution i ↓ ıknown, for sufficiently large sample sizes ($n > 30$), the process still holds due to the Central Limit Theorem.

2

1. What is a Confidence Interval?

A confidence interval (CI) provides a range of values within which we expect the true population parameter (like the mean) to fall, with a certain level of confidence (usually 95% or 90%).

---

2. Two Key Components of a Confidence Interval:

- **Sample Mean ($\bar{x}$):** The average of the data points in your sample.
- **Margin of Error (MOE):** This accounts for variability in your sample and the confidence level.

---

3. Formula for Confidence Interval:

$$\text{Confidence Interval} = \bar{x} \pm \text{Margin of Error}$$

Where:

- $\bar{x}$ is the sample mean.
- Margin of Error = $Z \times \frac{\sigma}{\sqrt{n}}$

---

4. Breaking Down the Margin of Error Formula:

- **Z-Score (Z):**
  This comes from the standard normal distribution and depends on your desired confidence level:
  ↓
  - 95% Confidence Level → $Z \approx 1.96$

3

4. Breaking Down the Margin of Error Formula:

- **Z-Score (Z):**
  This comes from the standard normal distribution and depends on your desired confidence level:
  - 95% Confidence Level → $Z \approx 1.96$
  - 90% Confidence Level → $Z \approx 1.65$
    The Z-score tells how many standard deviations from the mean capture the desired percentage of data.
- **Standard Deviation (σ):**
  The spread of the population data. If the population standard deviation is unknown, we estimate it using the sample standard deviation.
- **Sample Size (n):**
  The number of observations in the sample.

---

5. Steps to Construct the Confidence Interval:

Step 1: Compute the Sample Mean ($\bar{x}$)

Suppose you are measuring the height of people. Let's say your sample of 10 people gives an average height of 170 cm.

Step 2: Find the Z-Score

Choose the confidence level (say 95%). The corresponding Z-score is 1.96.

Step 3: Compute the Standard Error

The formula for the standard error (SE) is:

$$SE = \frac{\sigma}{}$$

4

Step 3: Compute the Standard Error

The formula for the standard error (SE) is:

$$SE = \frac{\sigma}{\sqrt{n}}$$

Assume the population standard deviation (σ) is 5 cm, and the sample size (n) is 10.

$$SE = \frac{5}{\sqrt{10}} \approx 1.58$$

Step 4: Calculate the Margin of Error (MOE)

$$MOE = Z \times SE = 1.96 \times 1.58 \approx 3.1$$

Step 5: Compute the Confidence Interval

$$CI = \bar{x} \pm MOE$$

$$CI = 170 \pm 3.1$$

The confidence interval is (166.9 cm, 173.1 cm).

---

6. Interpretation

"We are 95% confident that the true mean height of the population lies between 166.9 cm and 173.1 cm."

---

7. Key Assumptions for Using Z-Sc res:

- The population follows a normal distribution, or

5

7. Key Assumptions for Using Z-Scores:

- The population follows a normal distribution, or
- The sample size is large ($n \geq 30$) due to the Central Limit Theorem.

8. What if the Data is Not Normally Distributed?

If the population distribution is unknown or not normal, but the sample size is large, you can still use the above approach because of the **Central Limit Theorem**, which states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases.

9. Critical Points Recap:

- Z-scores come from the standard normal distribution.
- The standard error adjusts the standard deviation for the sample size.
- A larger sample size reduces the margin of error.
- The confidence interval provides a probabilistic boundary for the true population mean.

Would you like additional examples or visualization to make this clearer?

6

Why Does the Z-Score Have a Fixed Value?

The **Z-score** comes from the **standard normal distribution**, which is a special type of normal distribution with a mean of 0 and a standard deviation of 1.

The Z-score for a given confidence level is **pre-calculated** based on mathematical properties of this distribution. It tells us how many standard deviations away from the mean we need to go to capture a specific percentage of the data.

How is the Z-Score Derived?

The Z-score values come from **statistical tables (Z-tables)** or can be calculated using software (like Python or Excel).
For example, for a **95% confidence level**, the Z-score is **1.96**. This is because:

- The total area under the curve of a standard normal distribution is **1 (100%)**.
- In a 95% confidence interval, 95% of the data is in the middle, with **2.5% on each tail**.

A Z-score of 1.96 means:

> **"1.96 standard deviations from the mean captures 95% of the data."**

Why Can't We Compute the Z-Score Directly for Confidence Intervals?

Technically, **you don't need to compute it manually** because statisticians have already computed these values for standard confidence levels. You simply look them up.

However, if you want to compute it, here's h

1. For a **95% confidence level**, the left tail area is 0.025 (since 5% is divided into two tails)

7

Why Can't We Compute the Z-Score Directly for Confidence Intervals?

Technically, **you don't need to compute it manually** because statisticians have already computed these values for standard confidence levels. You simply look them up.

However, if you want to compute it, here's how:

1. For a **95% confidence level**, the left tail area is 0.025 (since 5% is divided into two tails).
2. You can use the inverse cumulative distribution function of a normal distribution.

In Python:

```python
from scipy.stats import norm
z_score = norm.ppf(0.975)  # 0.975 because 1 - 0.025 = 0.975
print(z_score)  # Outputs 1.959963984540054
```

This gives **1.96**, the familiar Z-score.

Why Use Fixed Z-Scores?

To save time and maintain accuracy, statisticians have defined **fixed Z-scores** for the most common confidence levels:

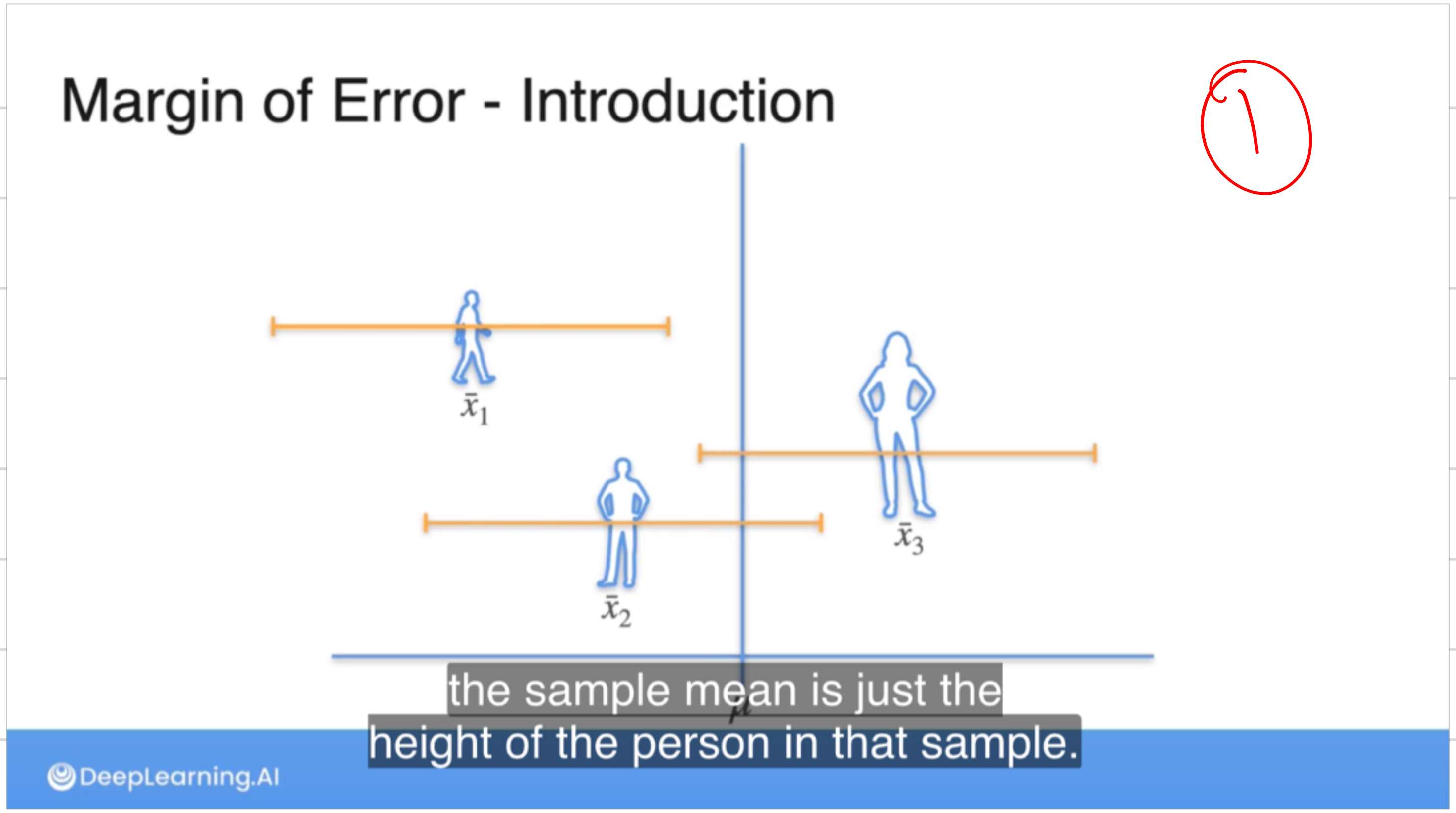
- **90% Confidence Level:** Z = 1.645
- **95% Confidence Level:** Z = 1.96
- **99% Confidence Level:** Z = 2.575

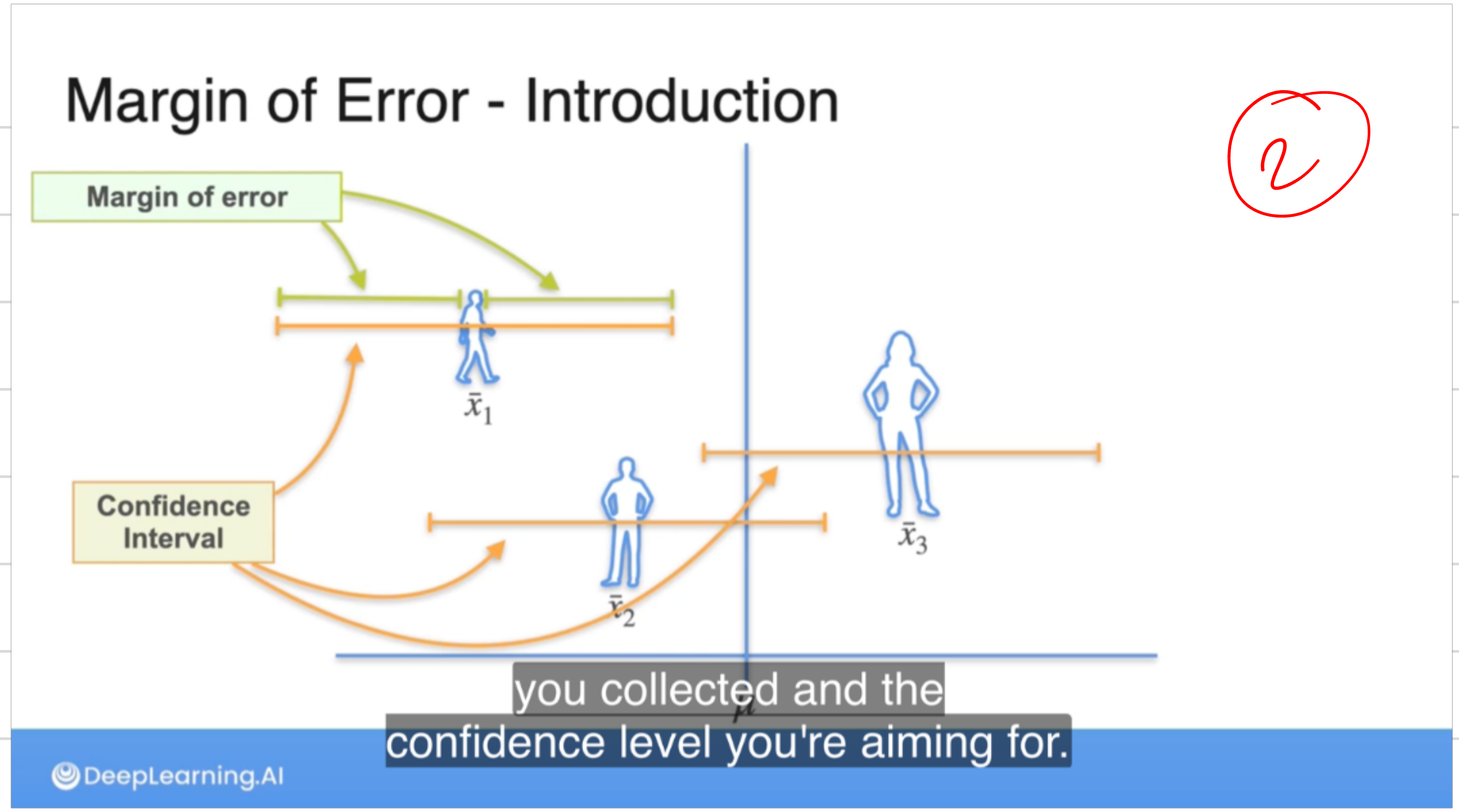
These fixed values simplify calculations without the need for recomputation every time.

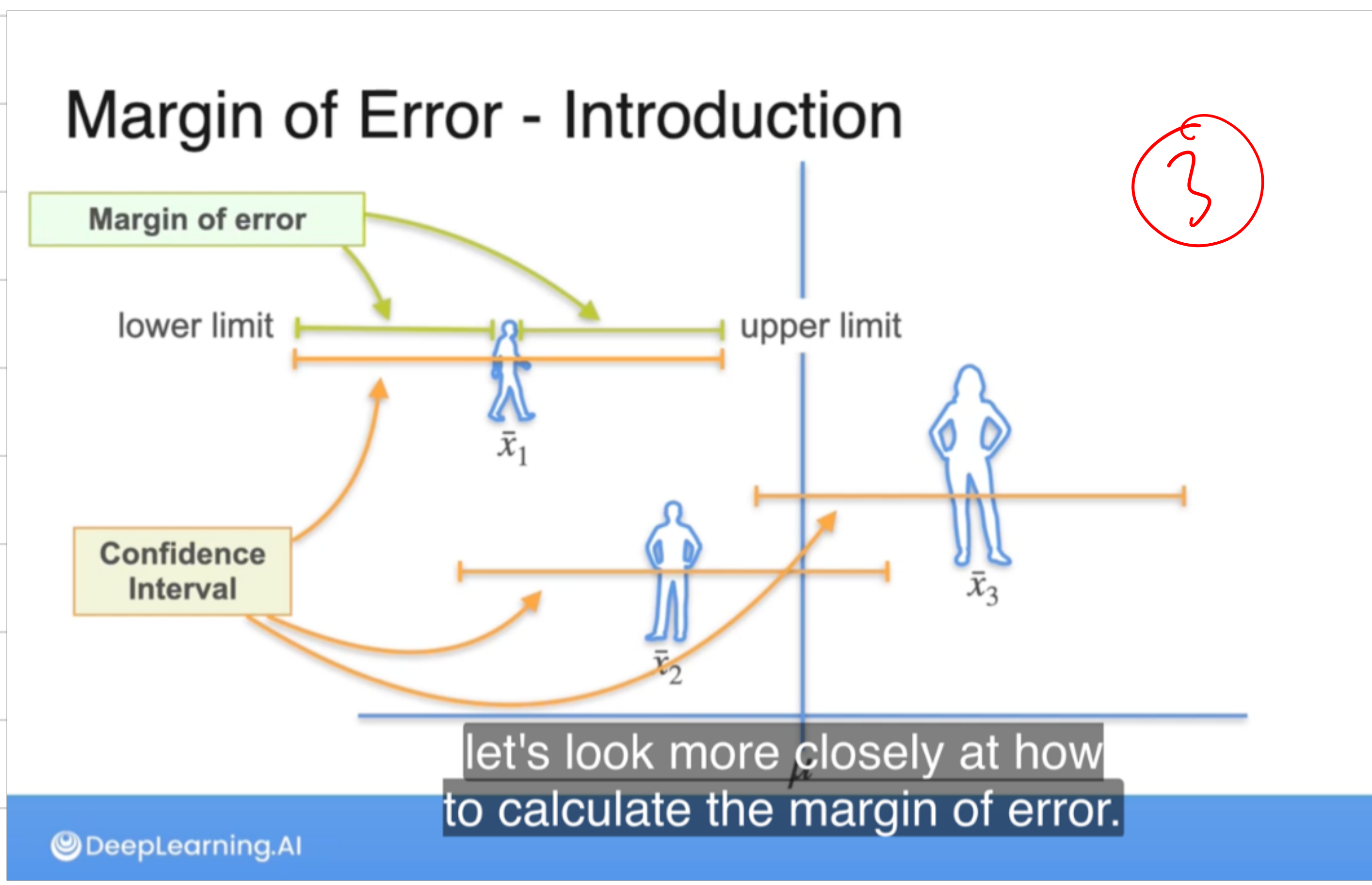
Significance level (α)

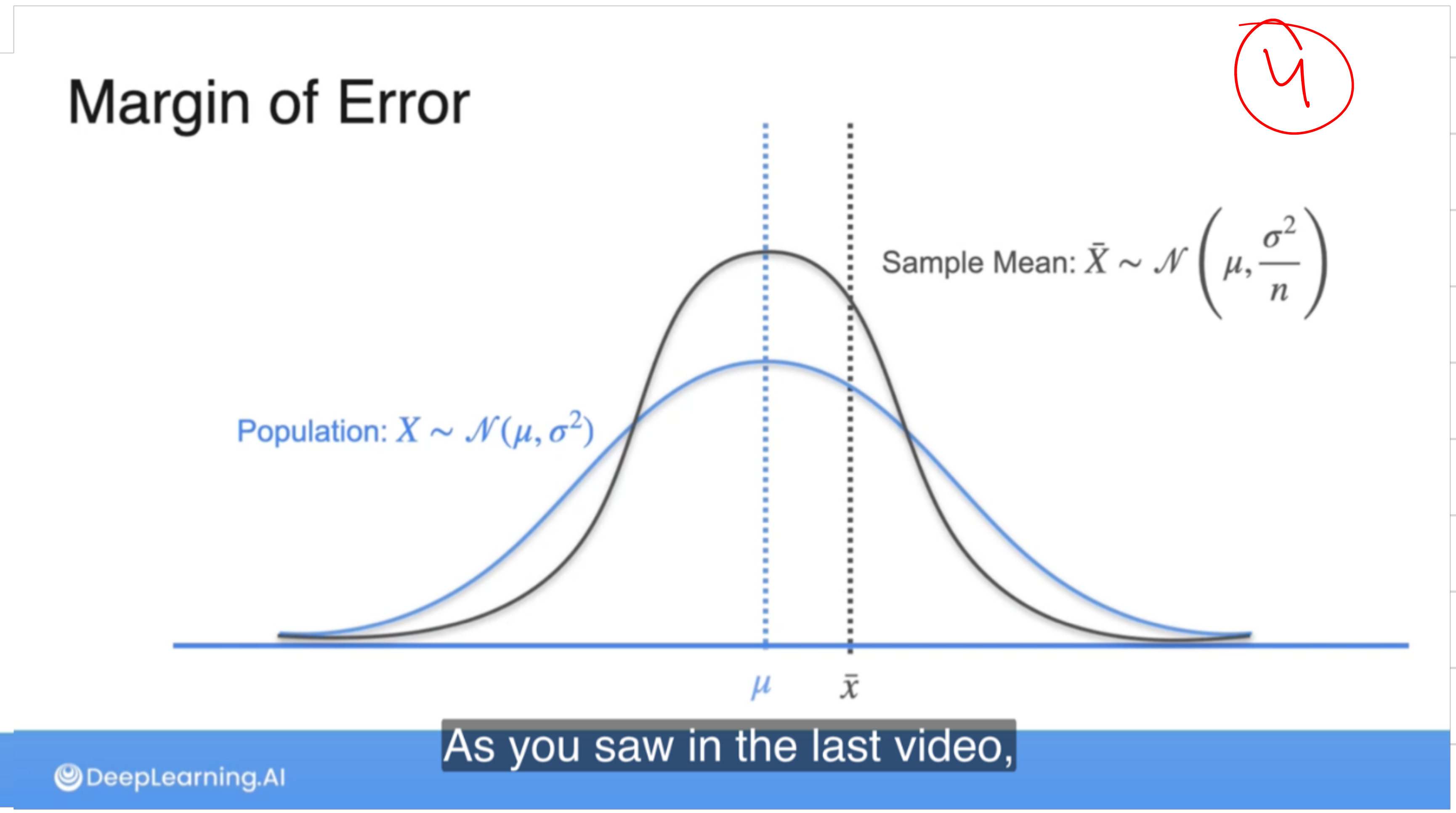
— 5% or 0.05 means we are willing to accept 5% chance that result happened by mistake.

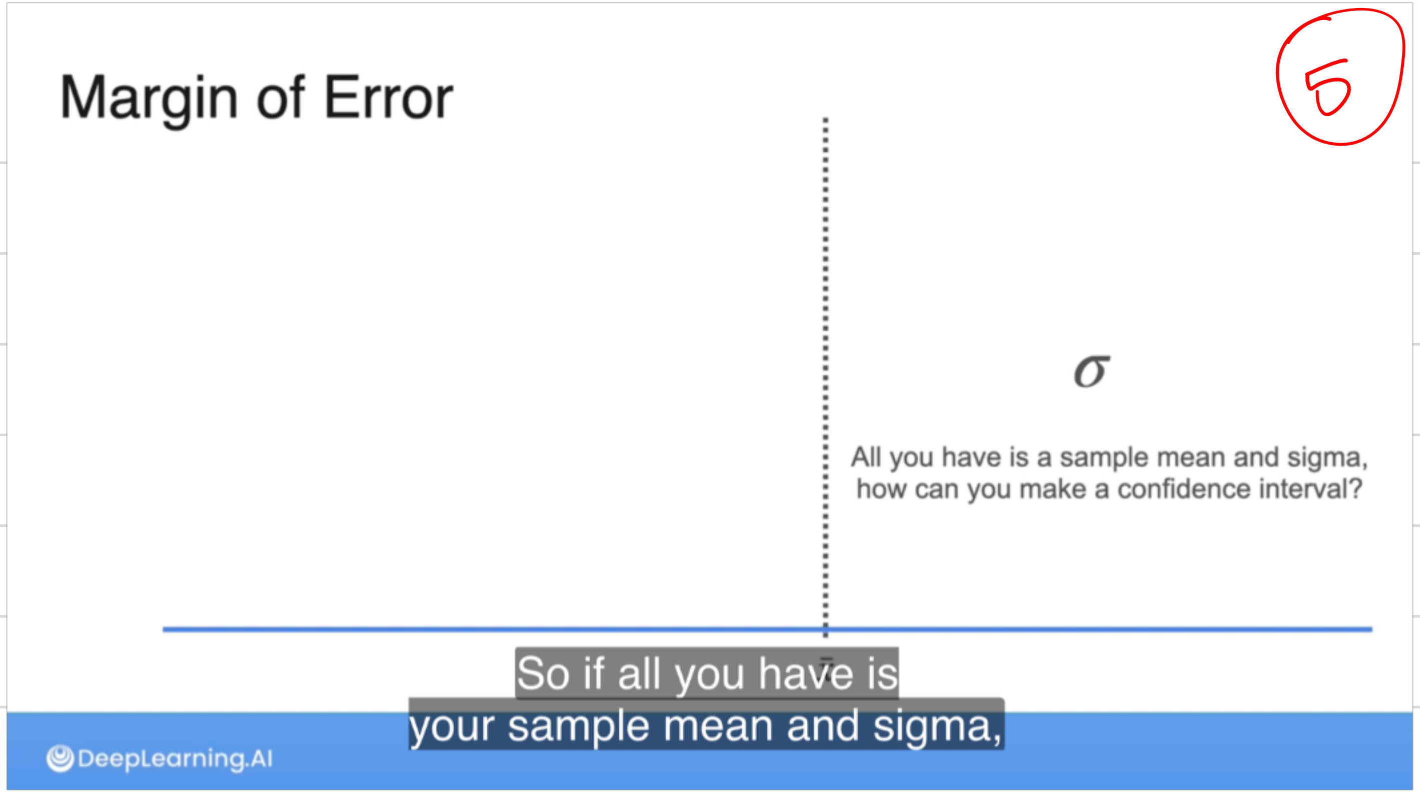
$\alpha = 0.05$ that Null Hypothesis the medicine does not work better. 'α' threshold used in statistical hypothesis to determine whether the null hypothesis can be rejected.

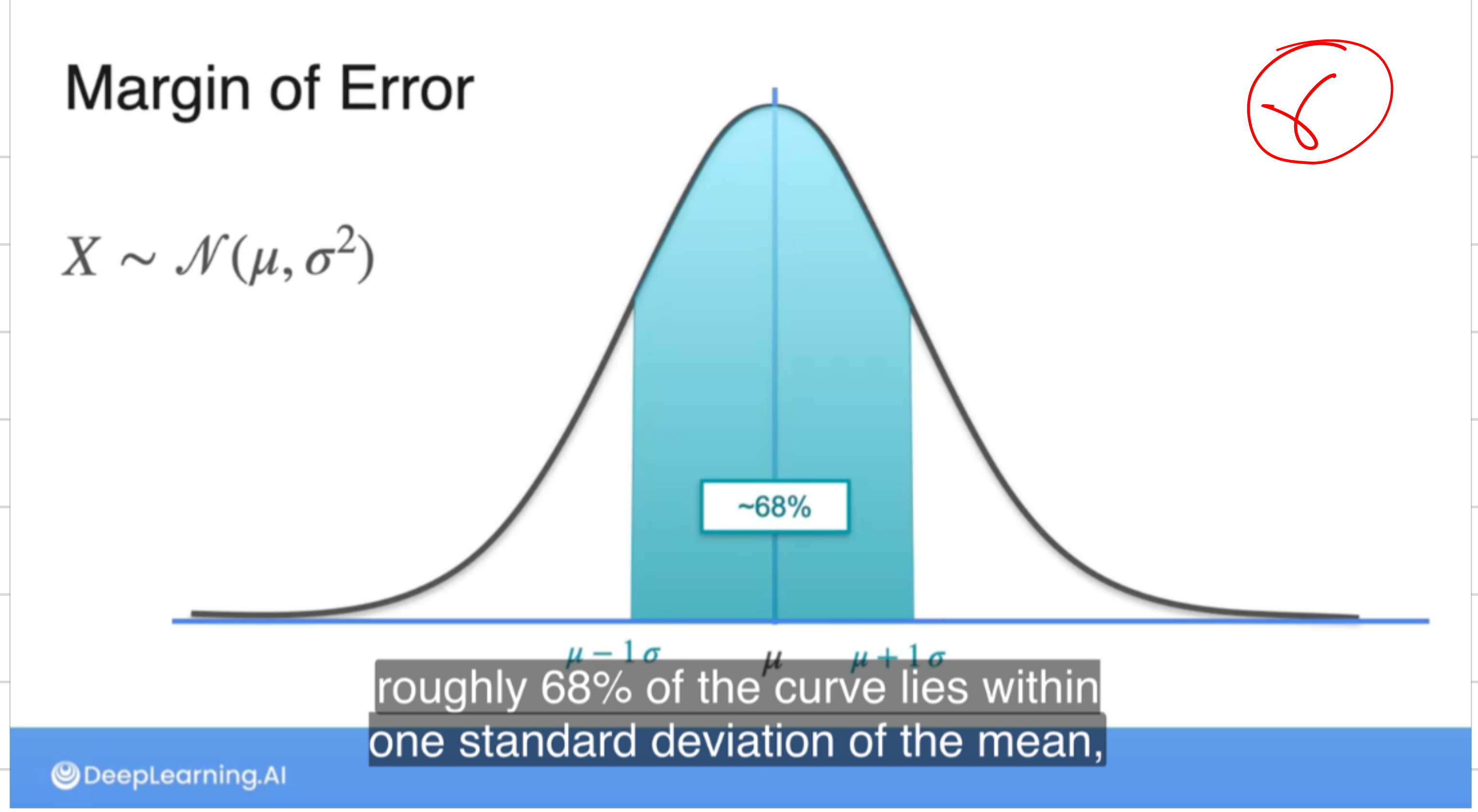

Margin of Error - Introduction
1
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
μ
the sample mean is just the
height of the person in that sample.
DeepLearning.AI

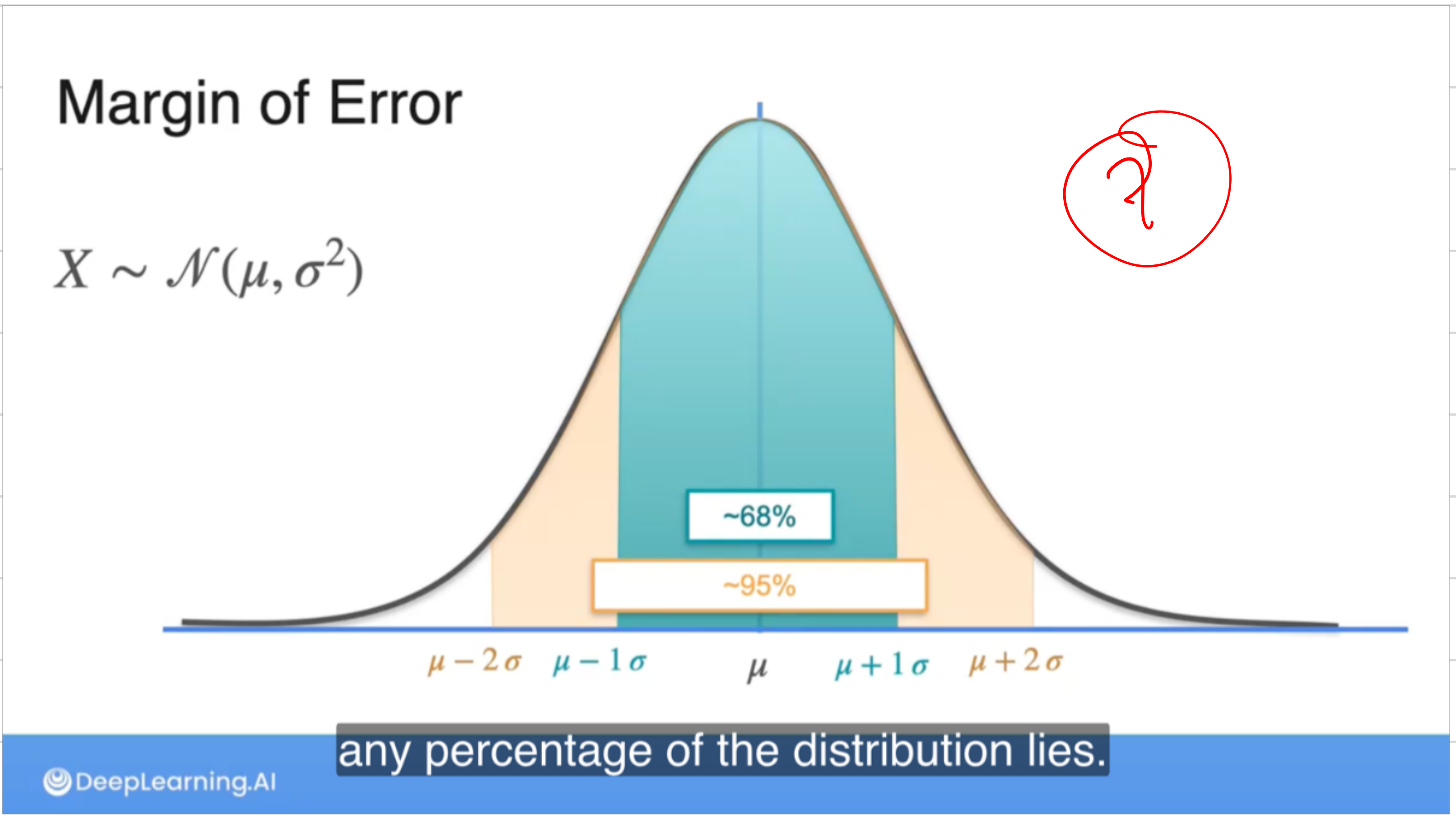

Margin of Error - Introduction
2
Margin of error
Confidence Interval
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
you collected and the
confidence level you're aiming for.
DeepLearning.AI

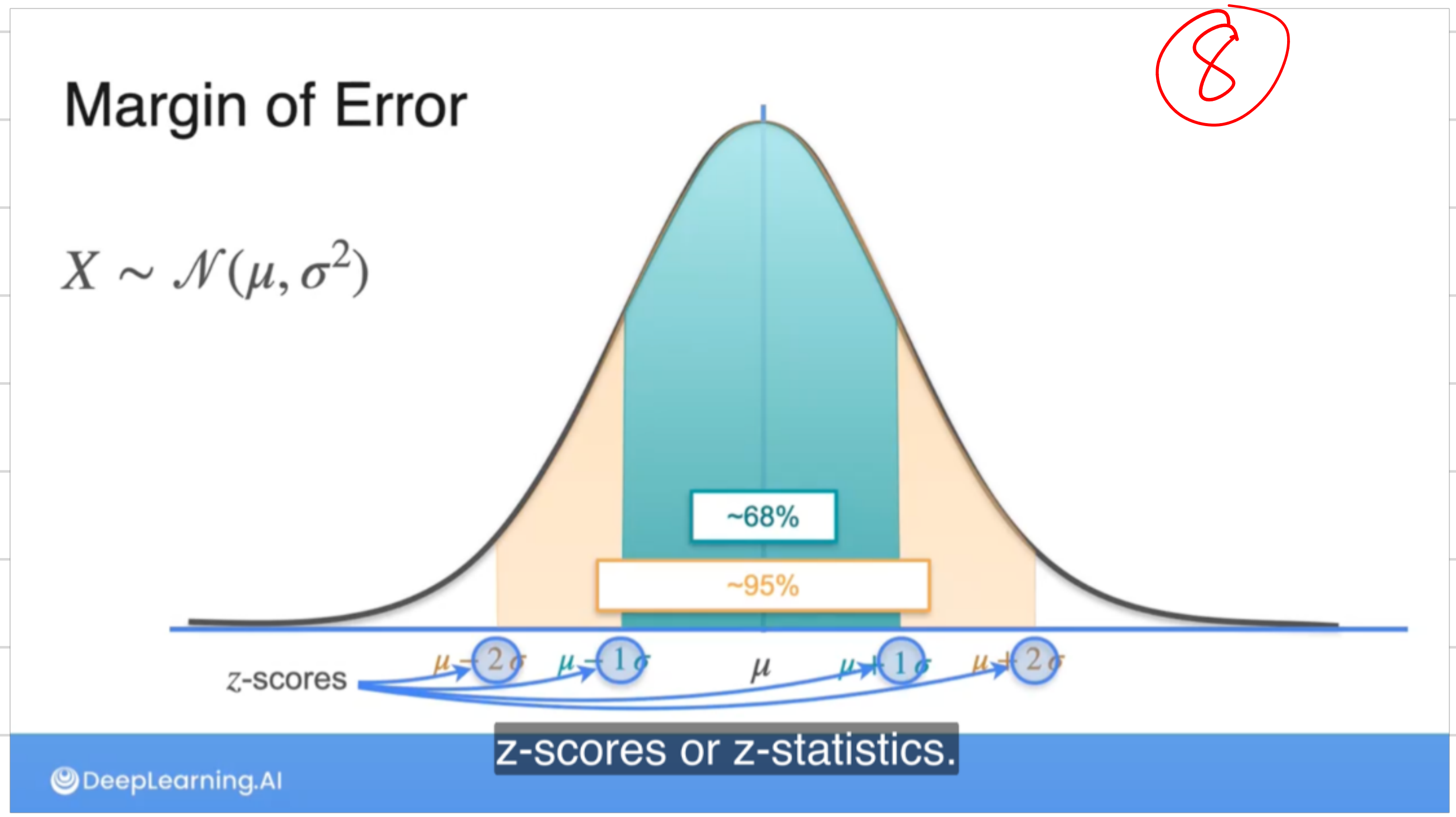

Margin of Error - Introduction
3
Margin of error
lower limit
upper limit
Confidence Interval
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
let's look more closely at how
to calculate the margin of error.
DeepLearning.AI

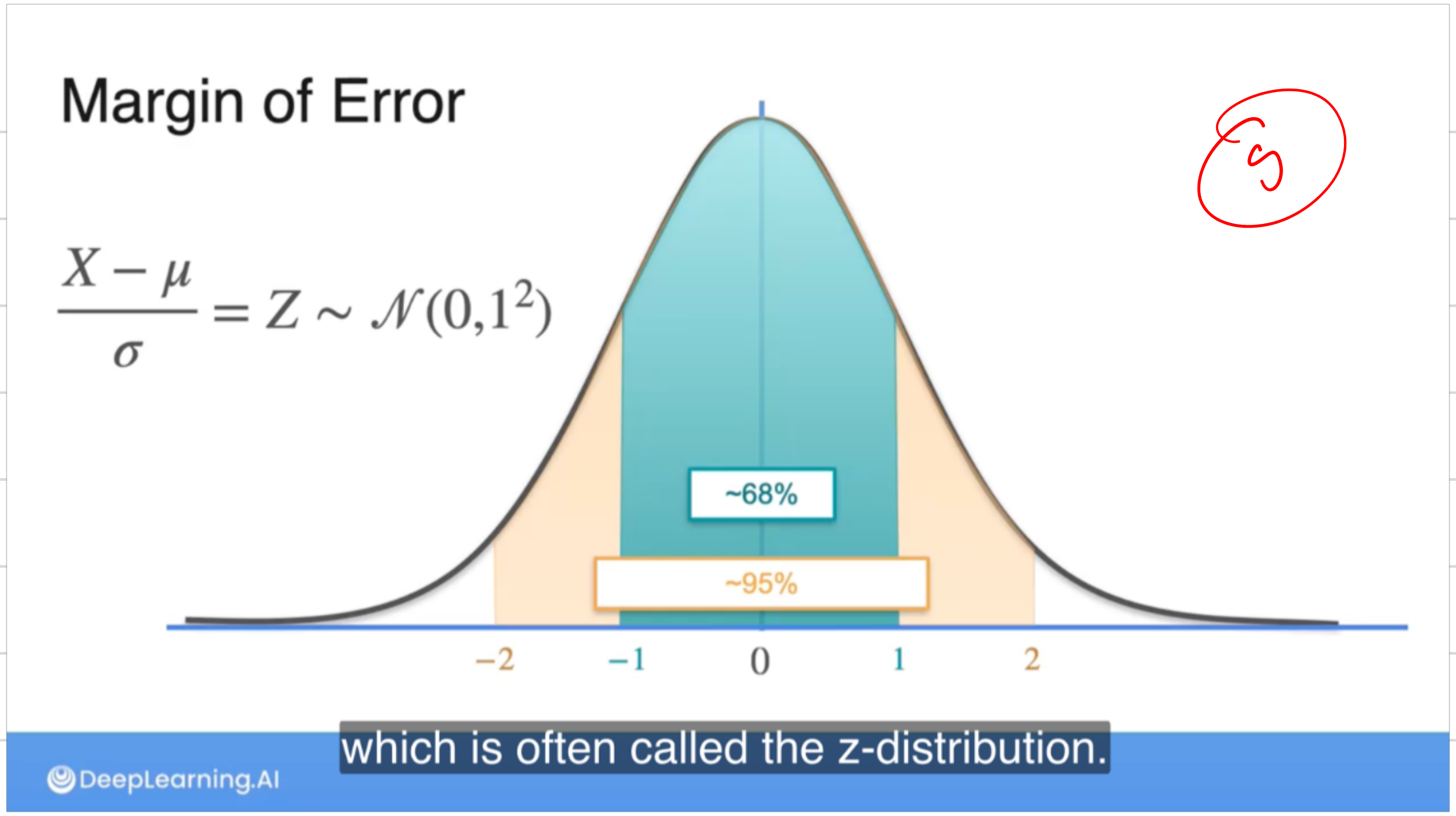

Margin of Error
4
Sample Mean: $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
Population: $X \sim \mathcal{N}(\mu, \sigma^2)$
μ
$\bar{x}$
As you saw in the last video,
DeepLearning.AI

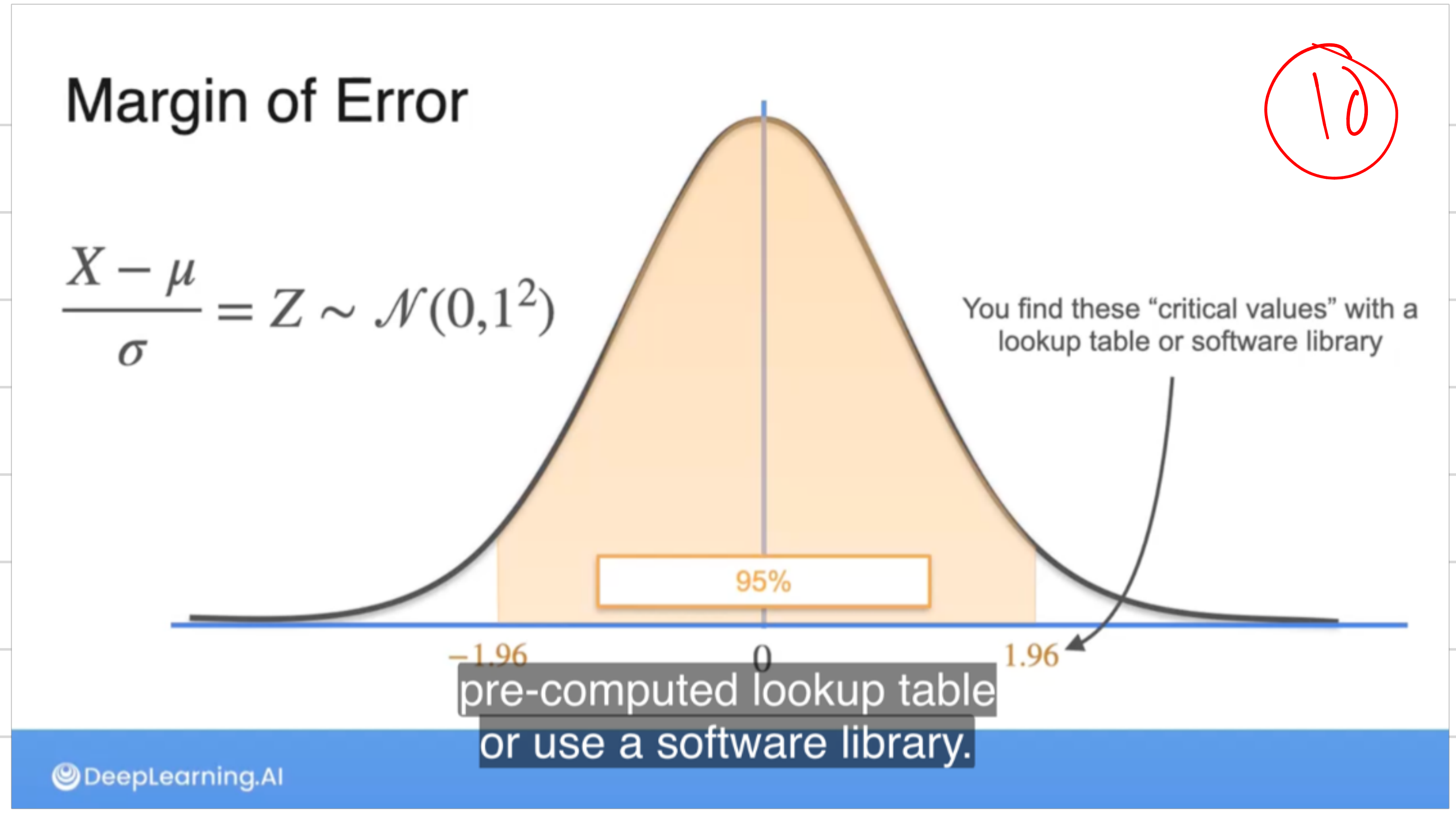

Margin of Error
5
σ
All you have is a sample mean and sigma,
how can you make a confidence interval?
So if all you have is
your sample mean and sigma,
DeepLearning.AI


Margin of Error
6
$X \sim \mathcal{N}(\mu, \sigma^2)$
~68%
$\mu - 1\sigma$
μ
$\mu + 1\sigma$
roughly 68% of the curve lies within
one standard deviation of the mean,
DeepLearning.AI


Margin of Error
7
$X \sim \mathcal{N}(\mu, \sigma^2)$
~68%
~95%
$\mu - 2\sigma$
$\mu - 1\sigma$
μ
$\mu + 1\sigma$
$\mu + 2\sigma$
any percentage of the distribution lies.
DeepLearning.AI


Margin of Error
8
$X \sim \mathcal{N}(\mu, \sigma^2)$
~68%
~95%
z-scores
μ
z-scores or z-statistics.
DeepLearning.AI





Question

11

When constructing a confidence interval at a 90% confidence level, which critical value should you use to calculate the margin of error?

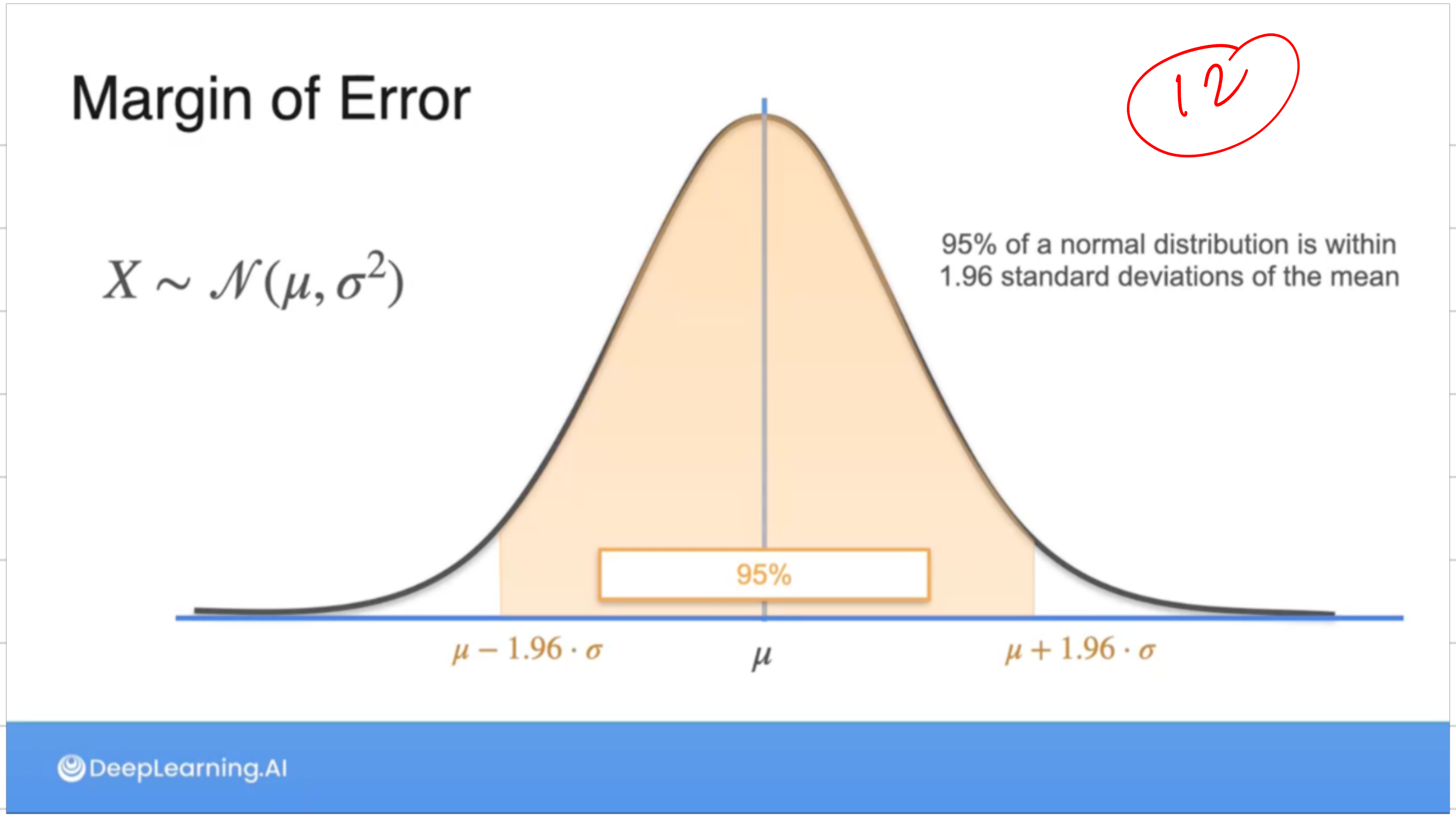
- ○ $z_{0.90}$
- ◉ $z_{0.95}$
- ○ $z_{0.025}$

Correct

Great job! When constructing a confidence interval at a 90% confidence level, you need to use the critical value associated with the significance level $\alpha = 0.10$. Therefore we look up the critical value at $z_{1-\alpha/2} = z_{1-0.10/2} = z_{0.05}$.

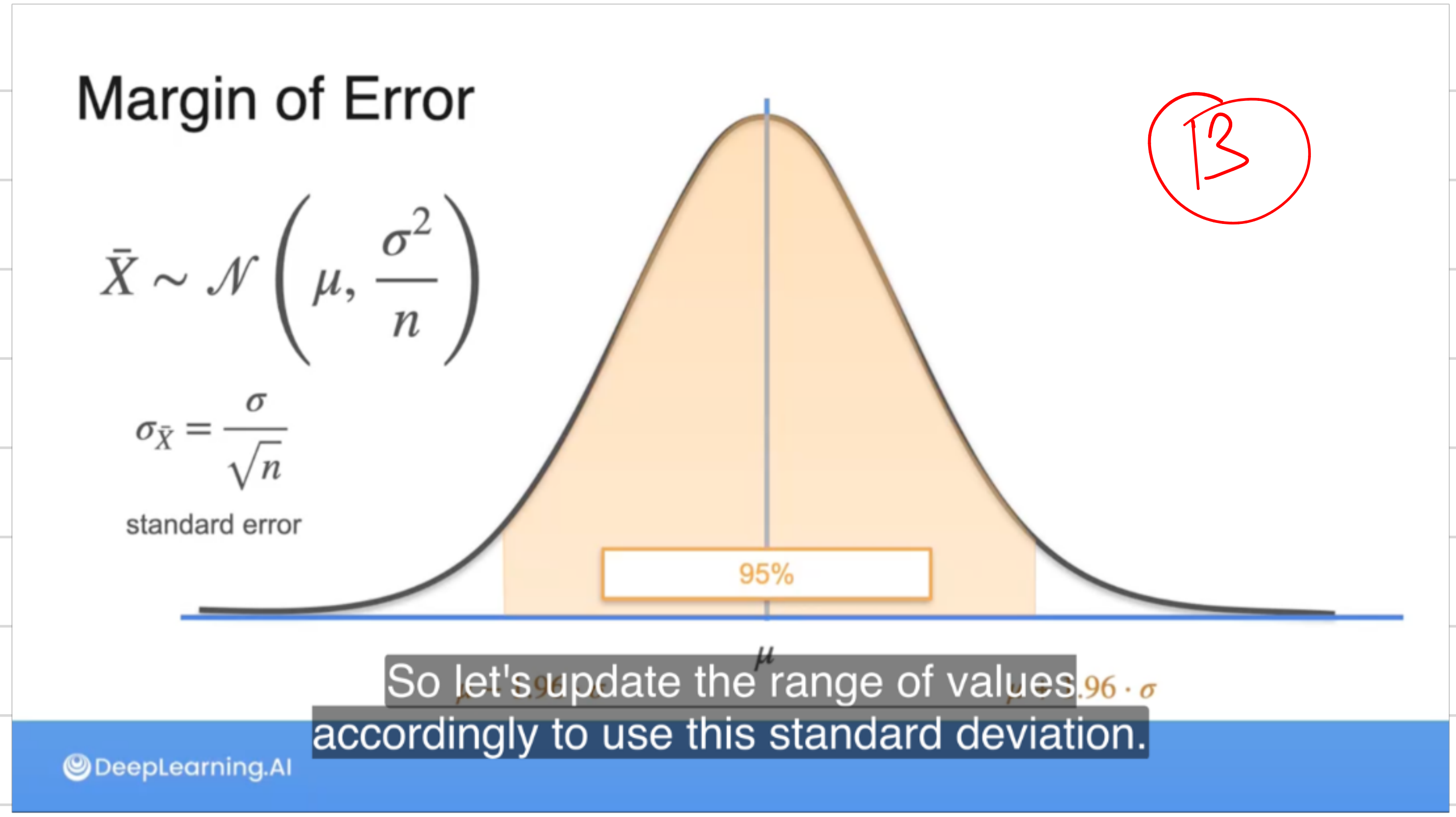
Skip Continue

12



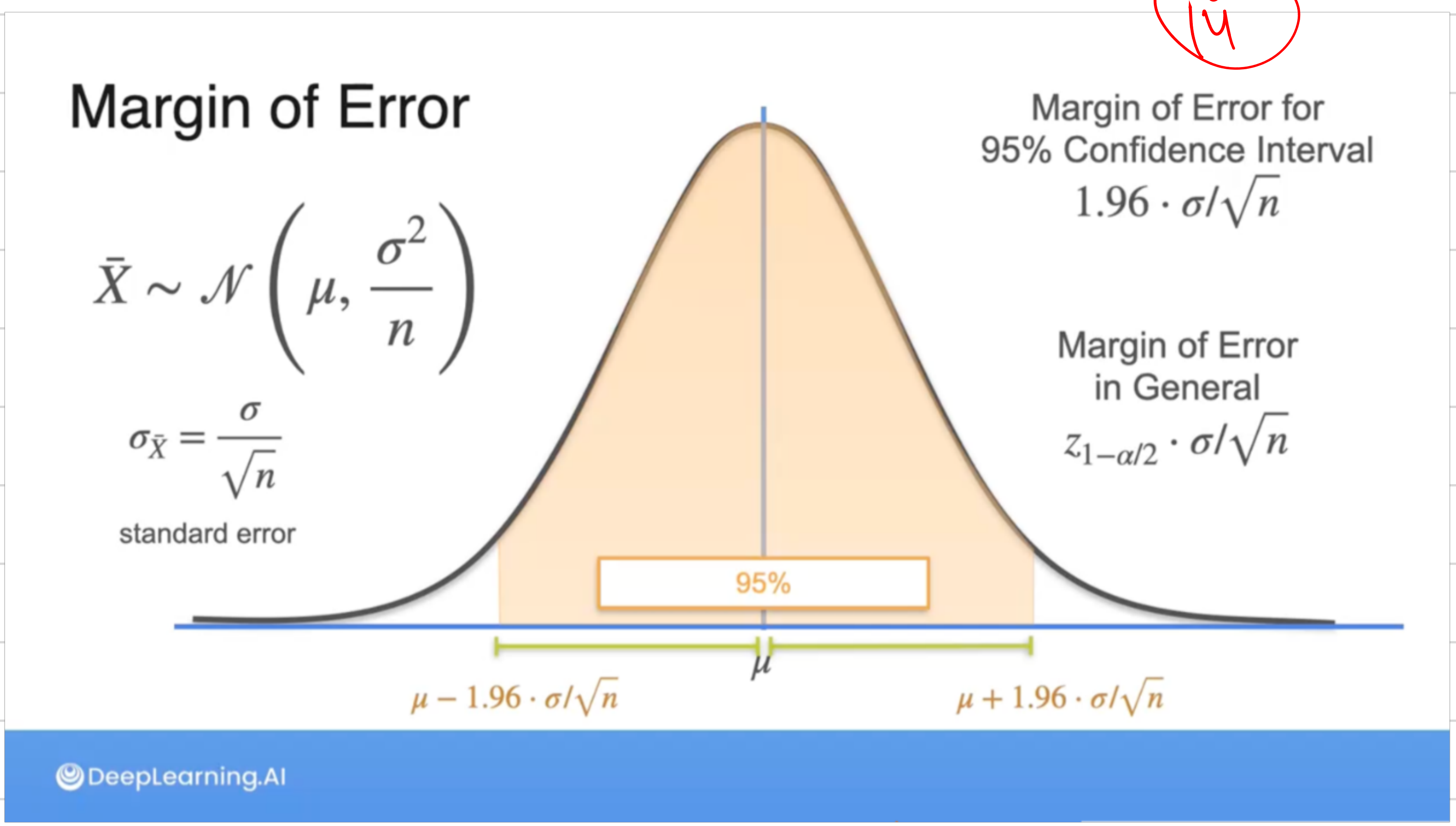
Margin of Error
$X \sim \mathcal{N}(\mu, \sigma^2)$
95% of a normal distribution is within
1.96 standard deviations of the mean
95%
$\mu - 1.96 \cdot \sigma$
μ
$\mu + 1.96 \cdot \sigma$
DeepLearning.AI

13



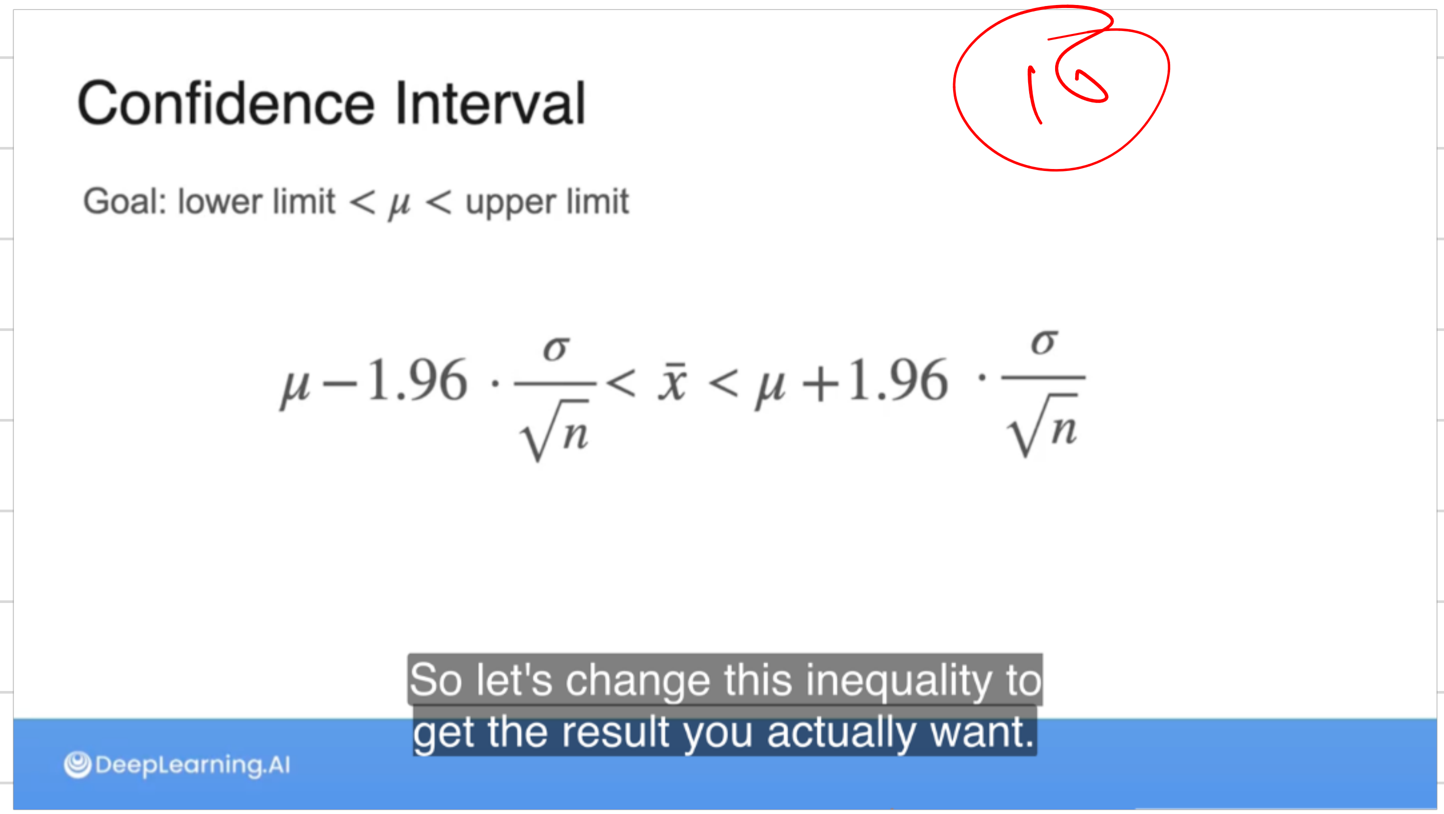
Margin of Error
$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
standard error
95%
μ
So let's update the range of values
accordingly to use this standard deviation.
DeepLearning.AI

14



Margin of Error
$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
standard error
Margin of Error for
95% Confidence Interval
$1.96 \cdot \sigma/\sqrt{n}$
Margin of Error
in General
$z_{1-\alpha/2} \cdot \sigma/\sqrt{n}$
95%
μ
$\mu - 1.96 \cdot \sigma/\sqrt{n}$
$\mu + 1.96 \cdot \sigma/\sqrt{n}$
DeepLearning.AI

15



Confidence Interval
Goal: lower limit $< \mu <$ upper limit
$\mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$
So let's change this inequality to
get the result you actually want.
DeepLearning.AI

18

Confidence Interval

Goal: lower limit $< \mu <$ upper limit

$$\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

confidence interval: $\bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

And so your confidence interval
ends up being the interval obtained

DeepLearning.AI

17

On the Distribution of Data

Population: $X \sim \mathcal{N}(\mu, \sigma^2)$

sample size n

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Population: X unknown (or not Normal)

sample size n

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

CLT

DeepLearning.AI

Source: Coursera Probability & Statistics for Machine Learning & Data Science:

1. CI - Overview
2. Margin of Error
3. Z-Score
4. Interactive Tools: Confidence Intervals.