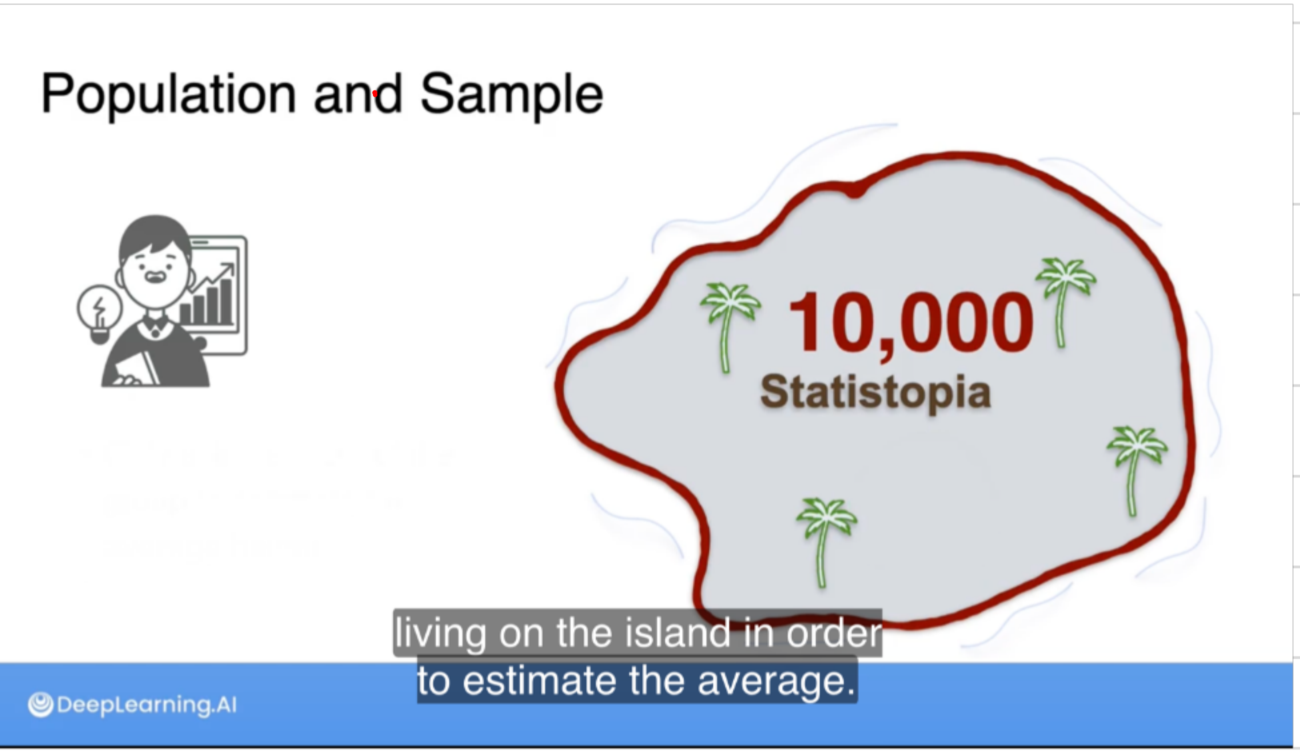
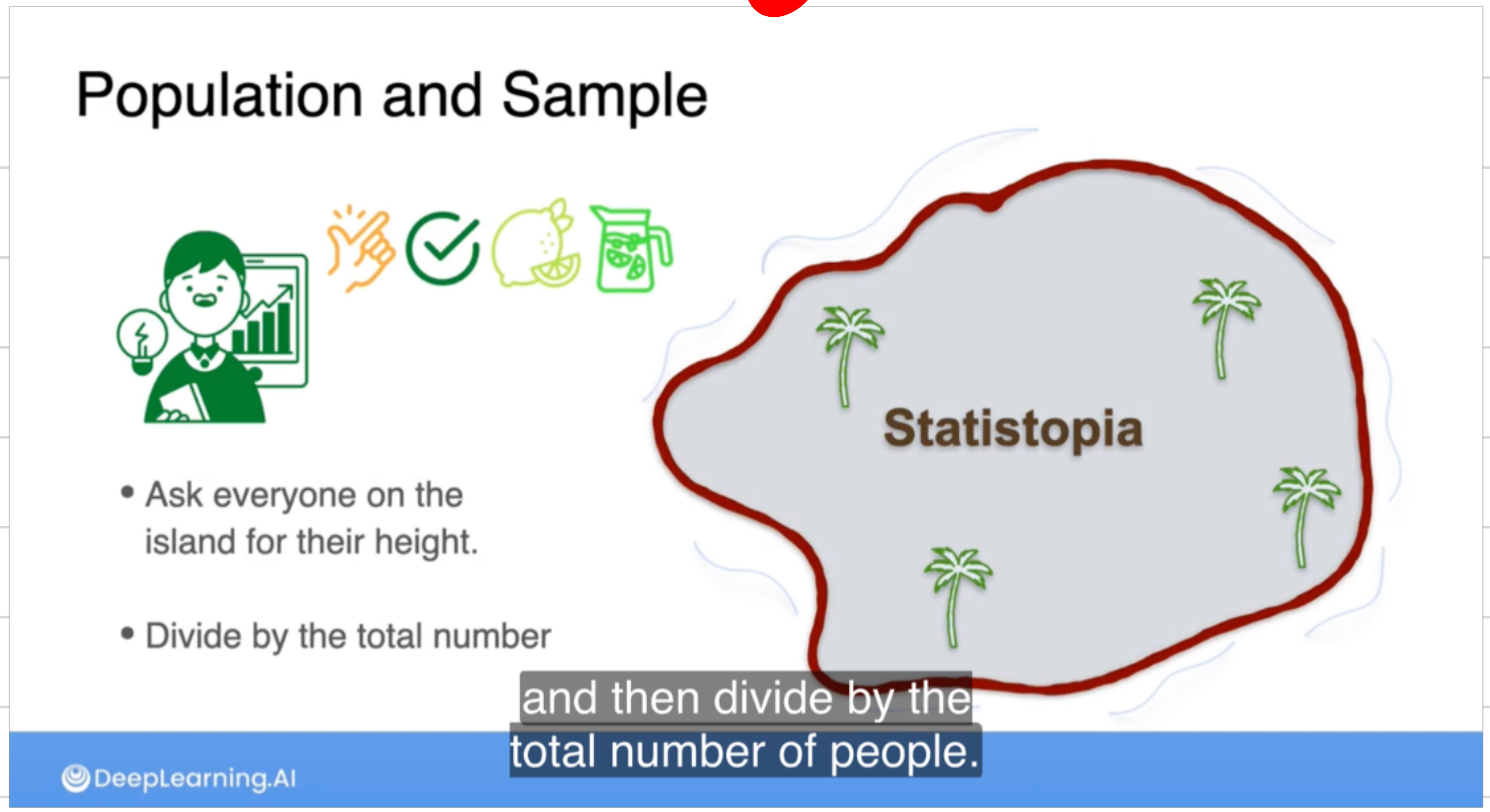
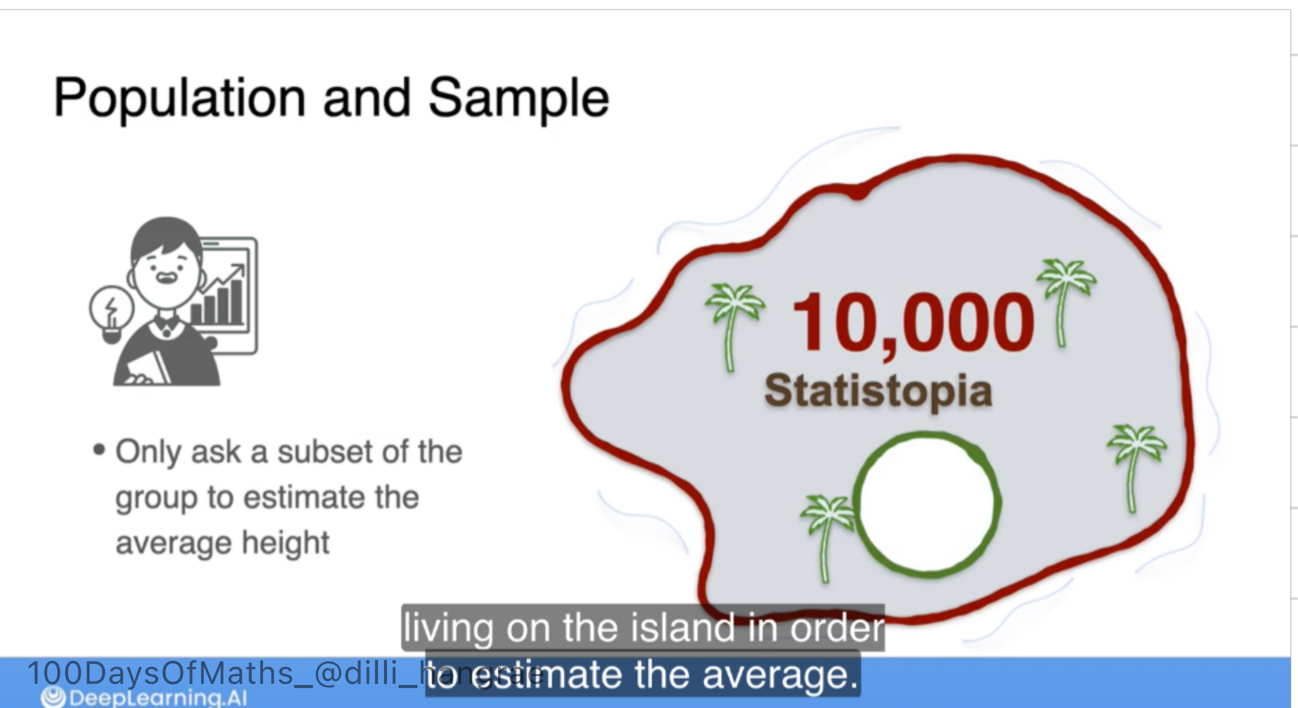
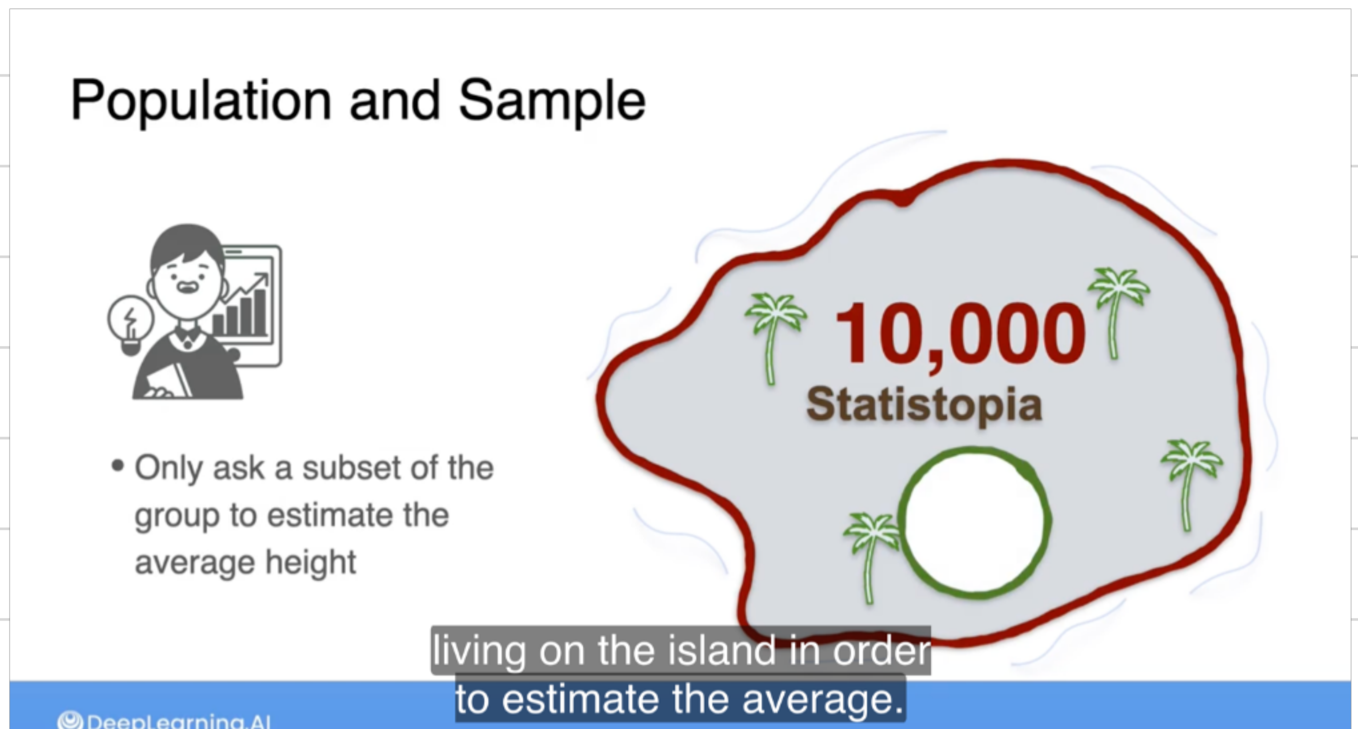
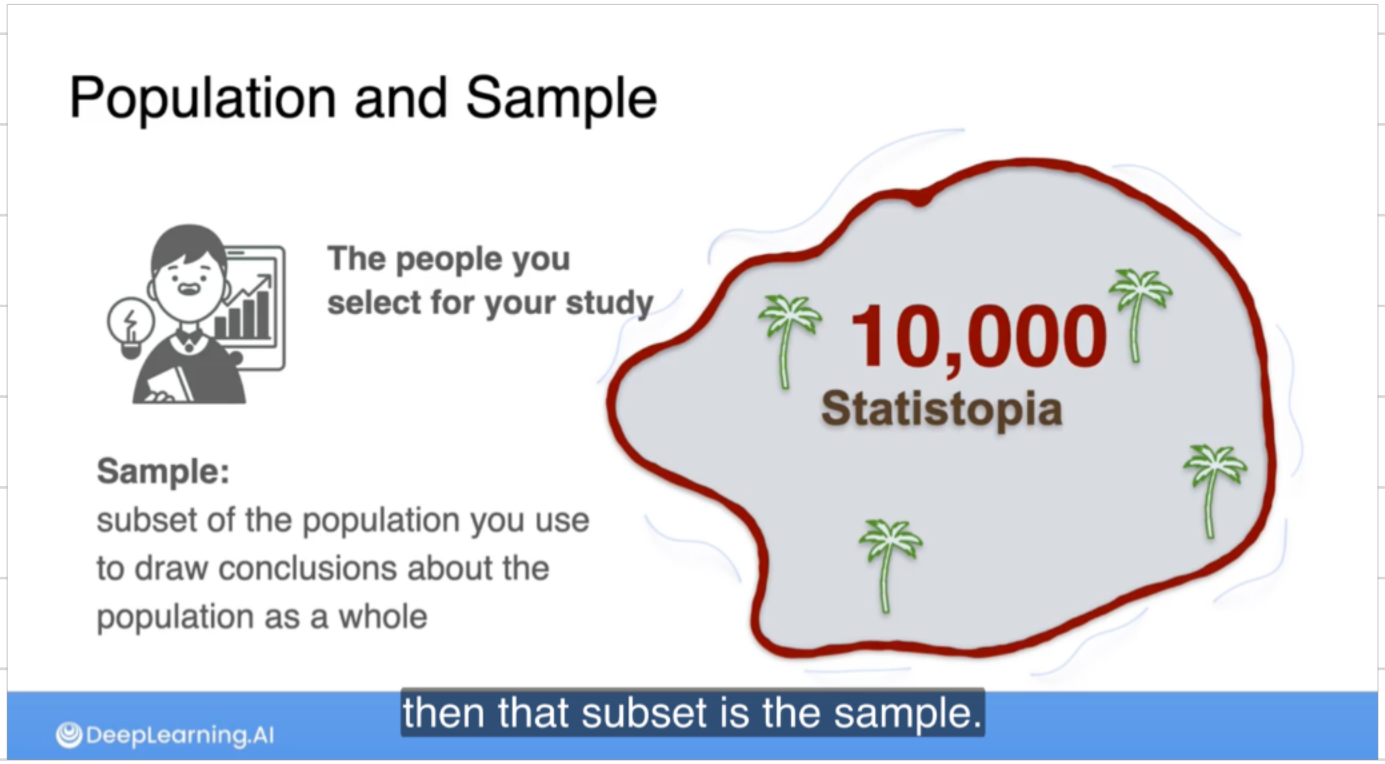
Day - 62, Jan - 31, 2025 (Magh 18, 2081)

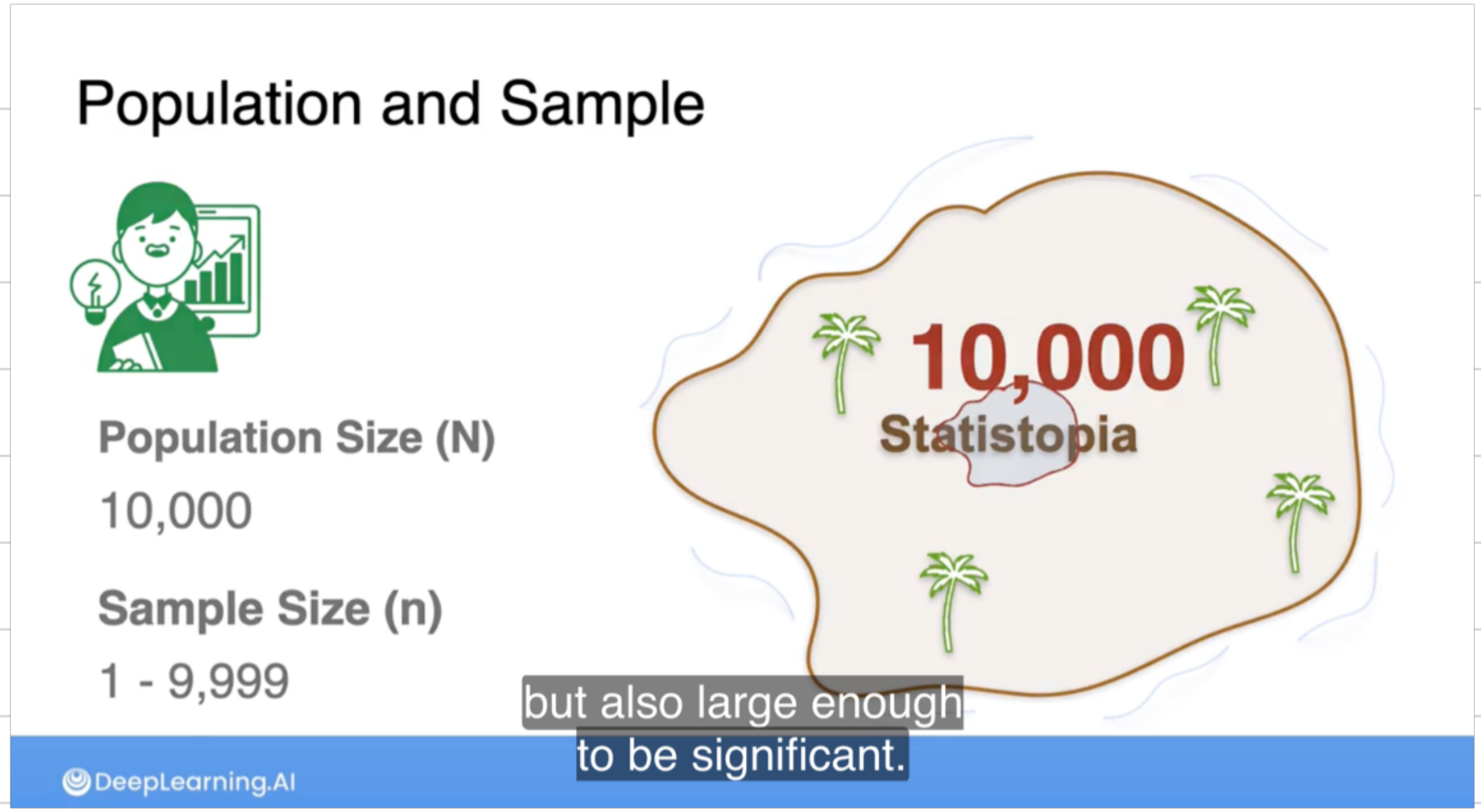


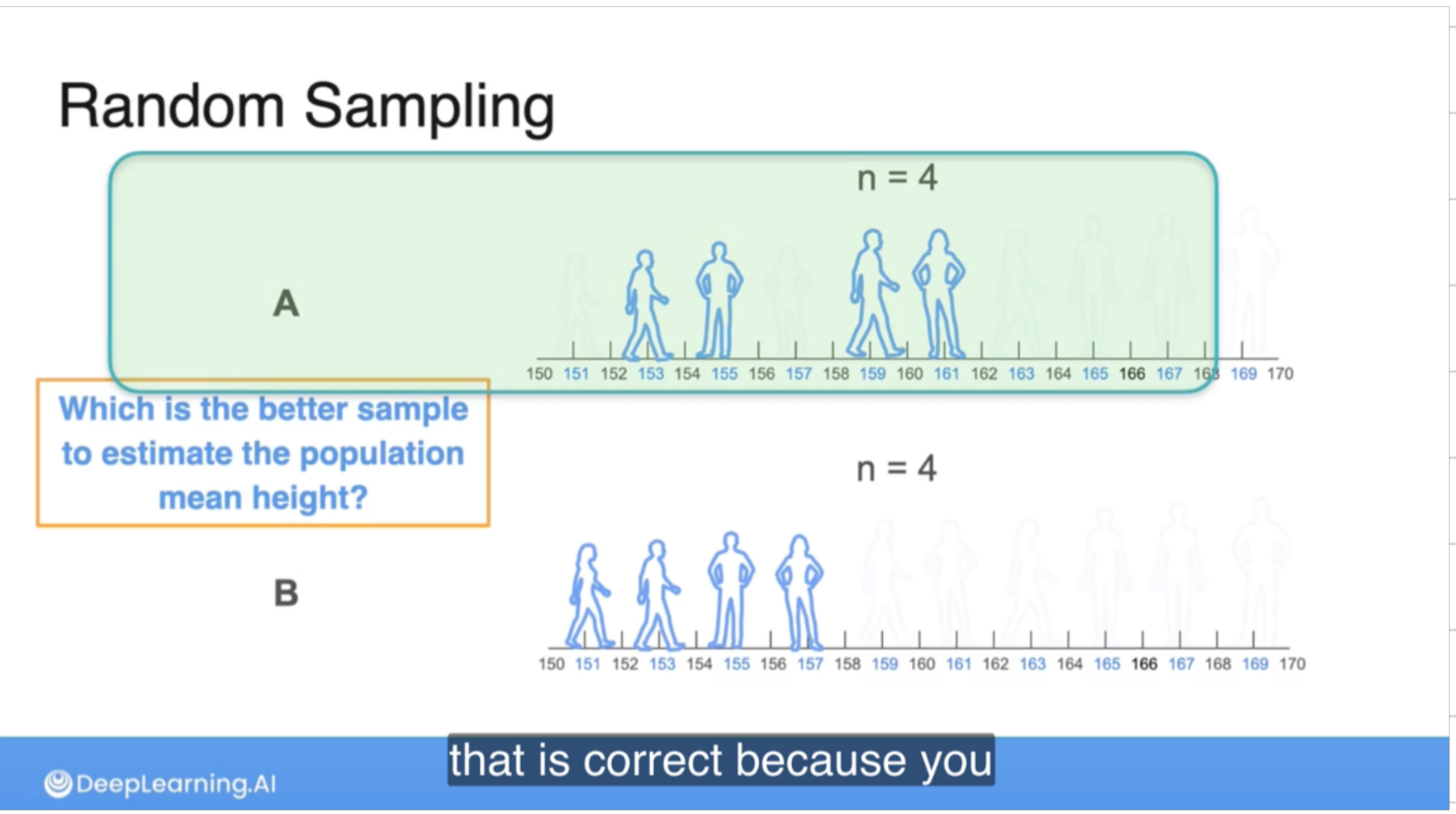


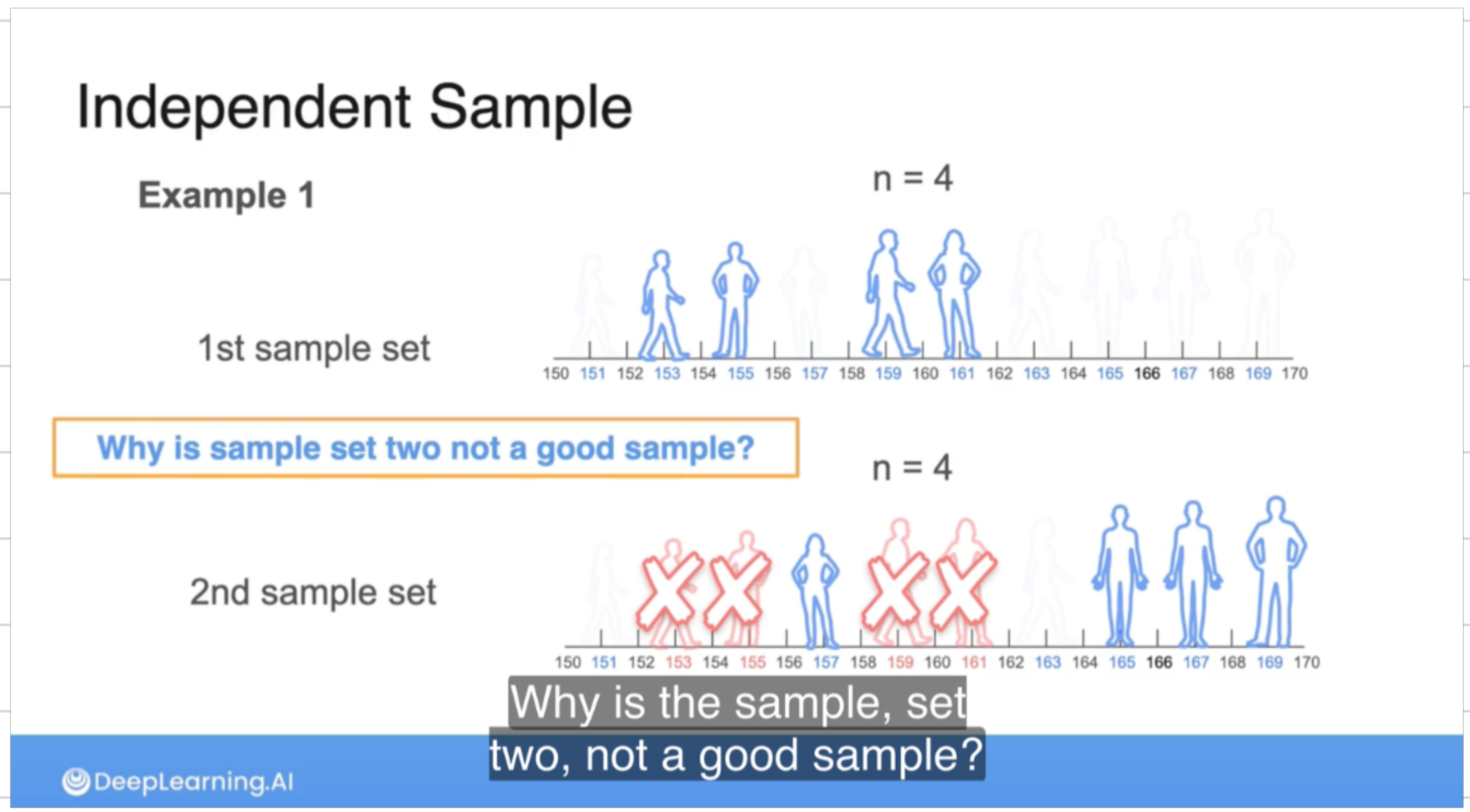


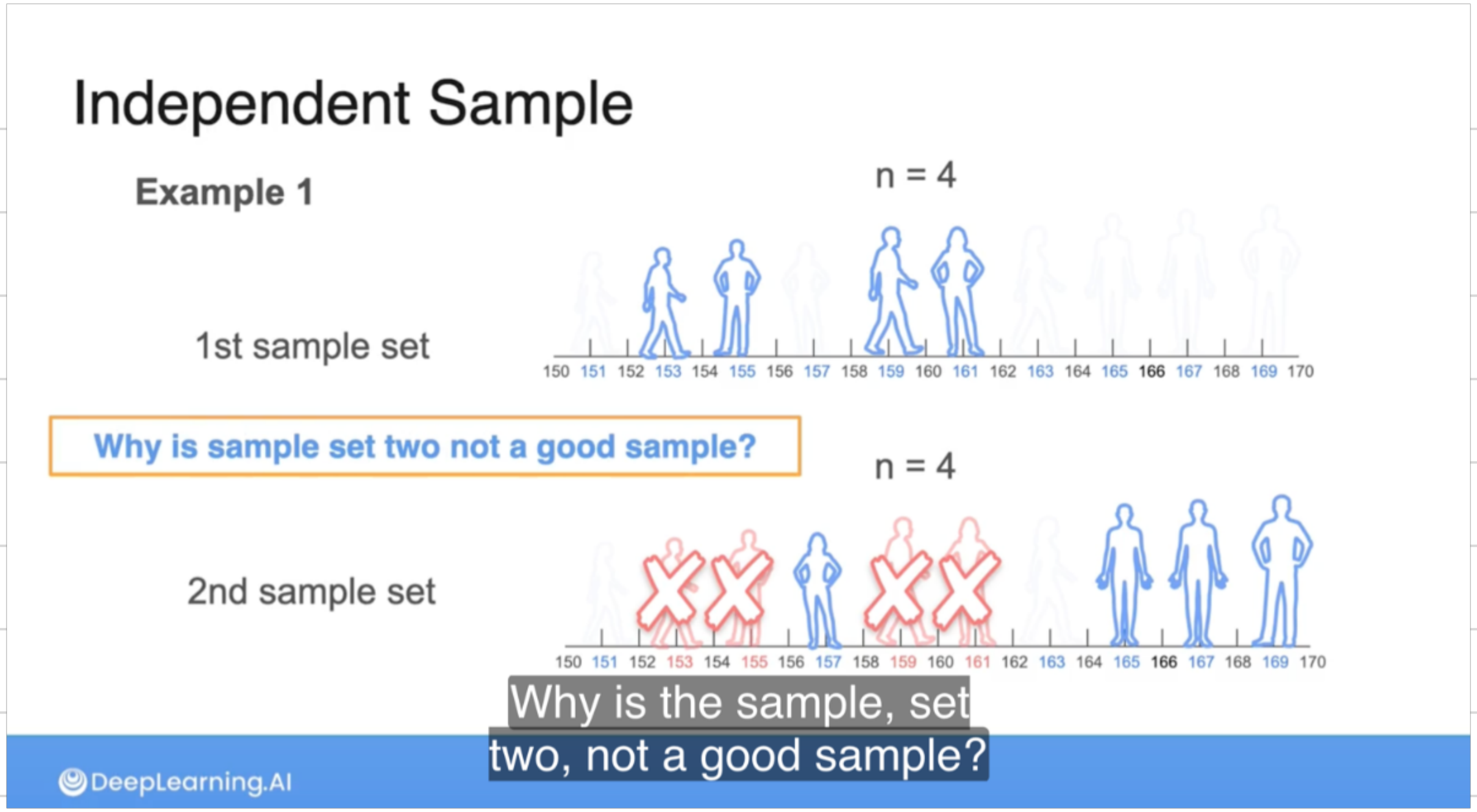


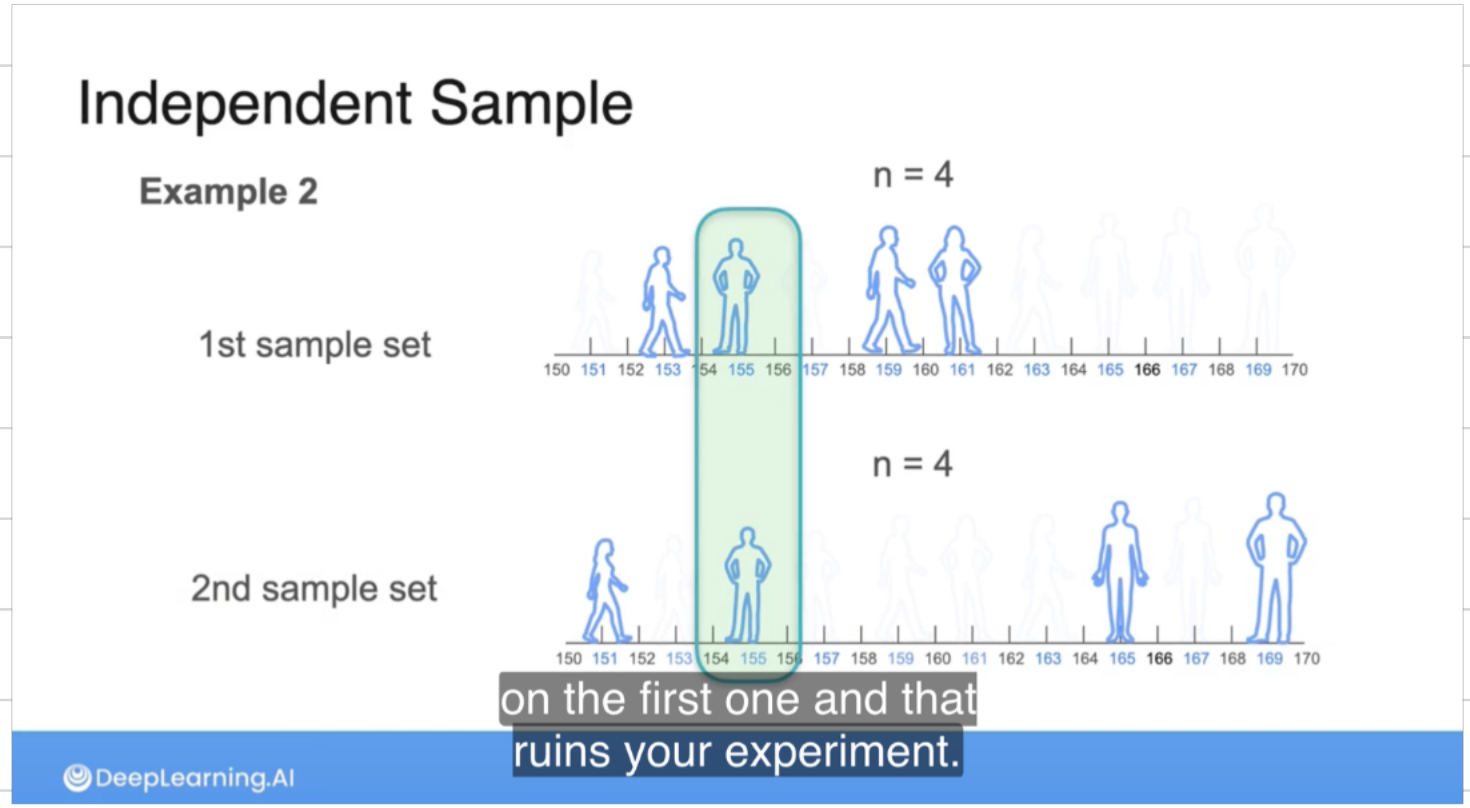

Population and Sample
The people you select for your study
10,000
Statistopia
Sample:
subset of the population you use to draw conclusions about the population as a whole
then that subset is the sample.
DeepLearning.AI

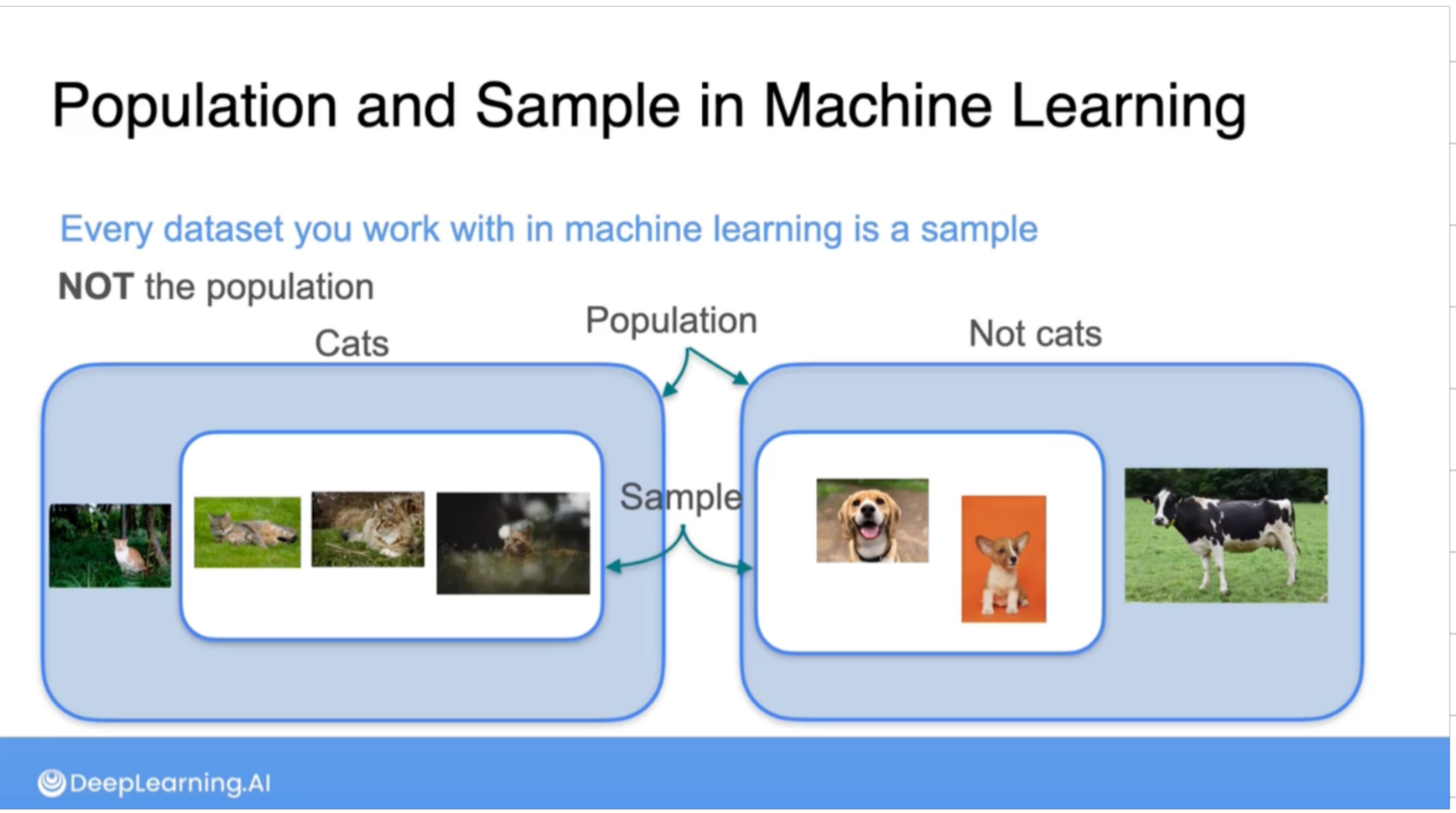

Population and Sample
Population Size (N)
10,000
Sample Size (n)
1 - 9,999
10,000
Statistopia
but also large enough
to be significant.
DeepLearning.AI

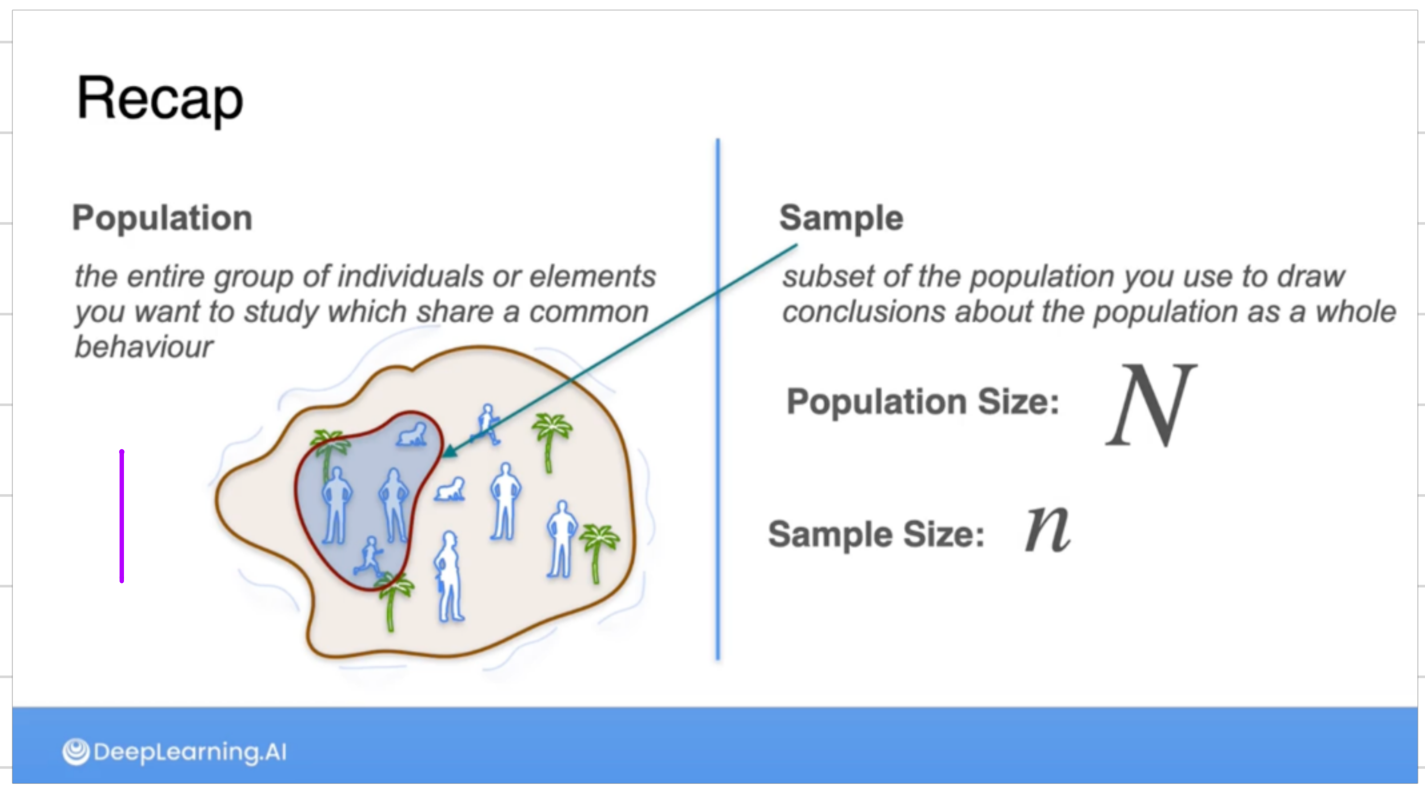

Random Sampling
n = 4
A
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
Which is the better sample to estimate the population mean height?
n = 4
B
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
that is correct because you
DeepLearning.AI


Independent Sample
Example 1
n = 4
1st sample set
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
Why is sample set two not a good sample?
n = 4
2nd sample set
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
Why is the sample, set
two, not a good sample?
DeepLearning.AI


Independent Sample
Example 1
n = 4
1st sample set
Why is sample set two not a good sample?
n = 4
2nd sample set
Why is the sample, set two, not a good sample?
DeepLearning.AI


Independent Sample
Example 2
n = 4
1st sample set
n = 4
2nd sample set
on the first one and that ruins your experiment.
DeepLearning.AI


Population and Sample in Machine Learning
Every dataset you work with in machine learning is a sample
NOT the population
Cats
Population
Not cats
Sample
DeepLearning.AI


Recap
Population
the entire group of individuals or elements you want to study which share a common behaviour
Sample
subset of the population you use to draw conclusions about the population as a whole
Population Size: N
Sample Size: n
DeepLearning.AI

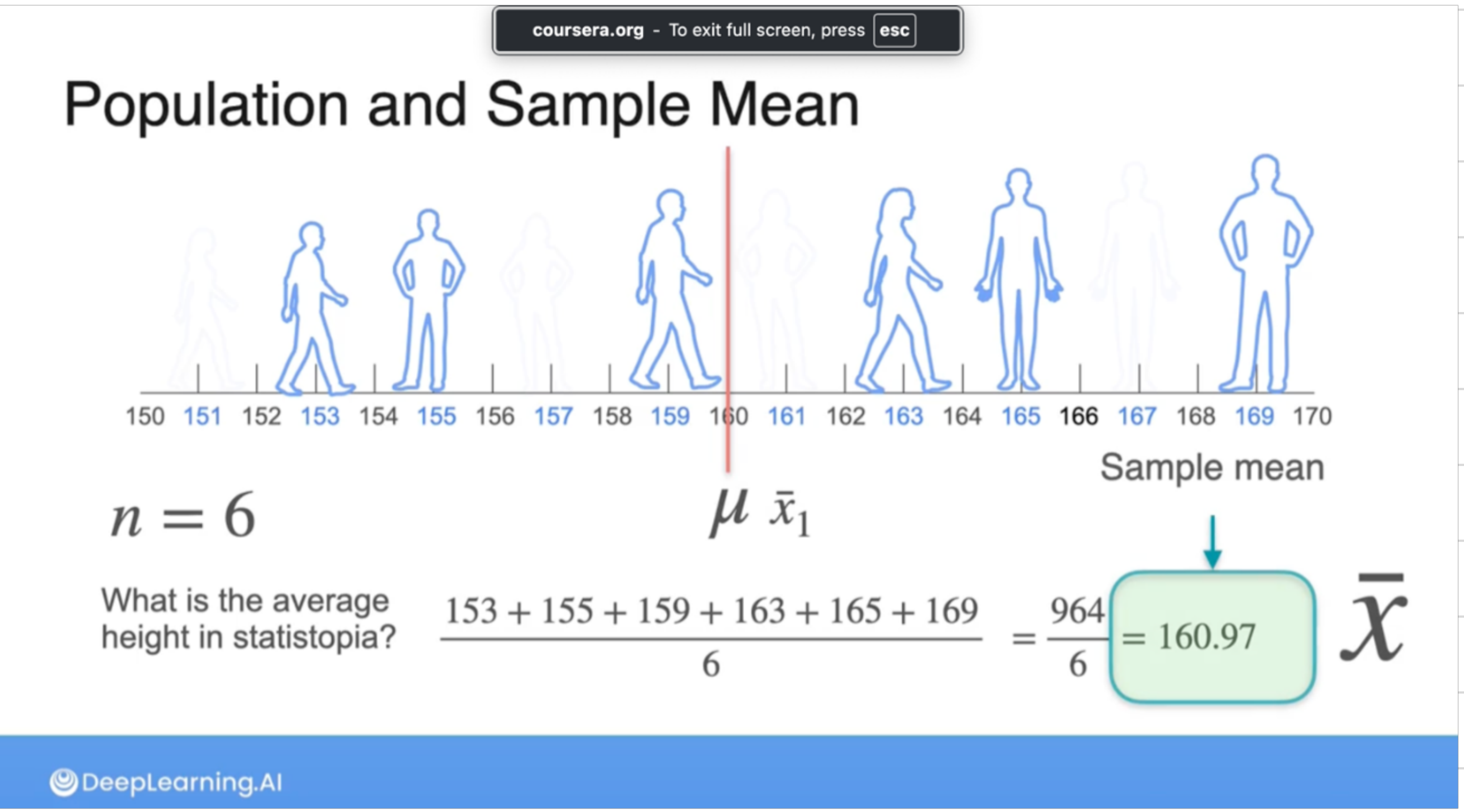
Which sample is most likely to generate a mean that is closest to the population mean?

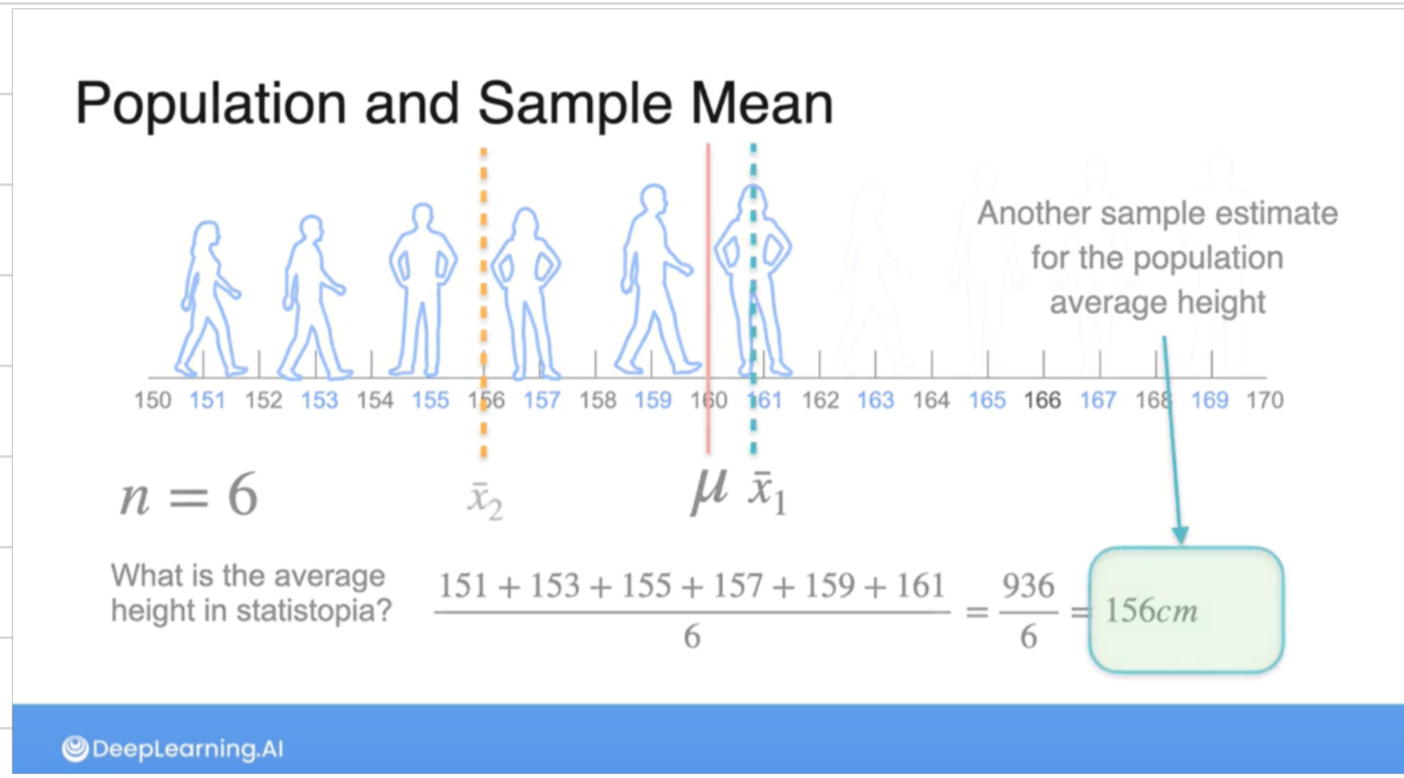
Random sample of 6 people

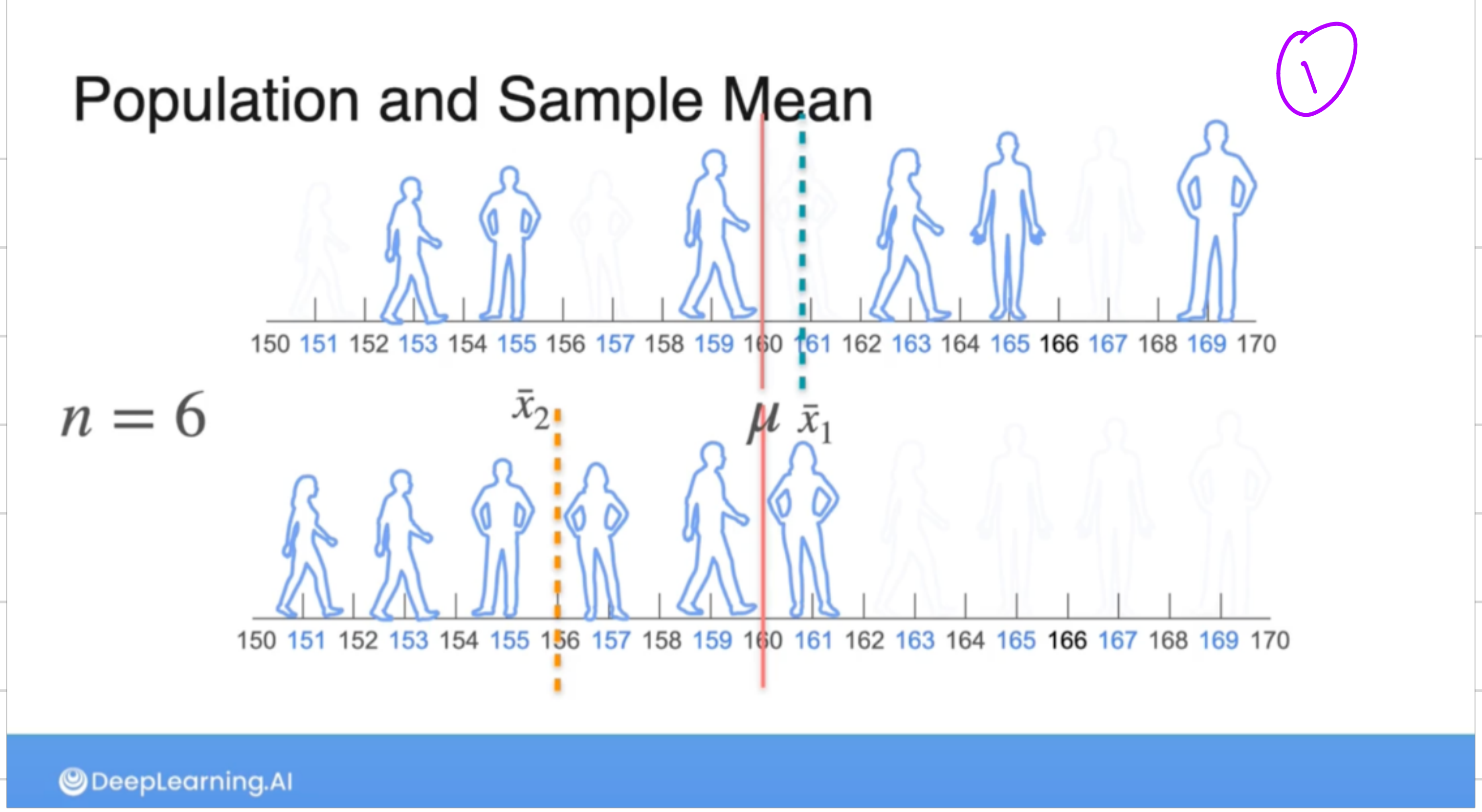
Convenience sample of the first 6 people observed.

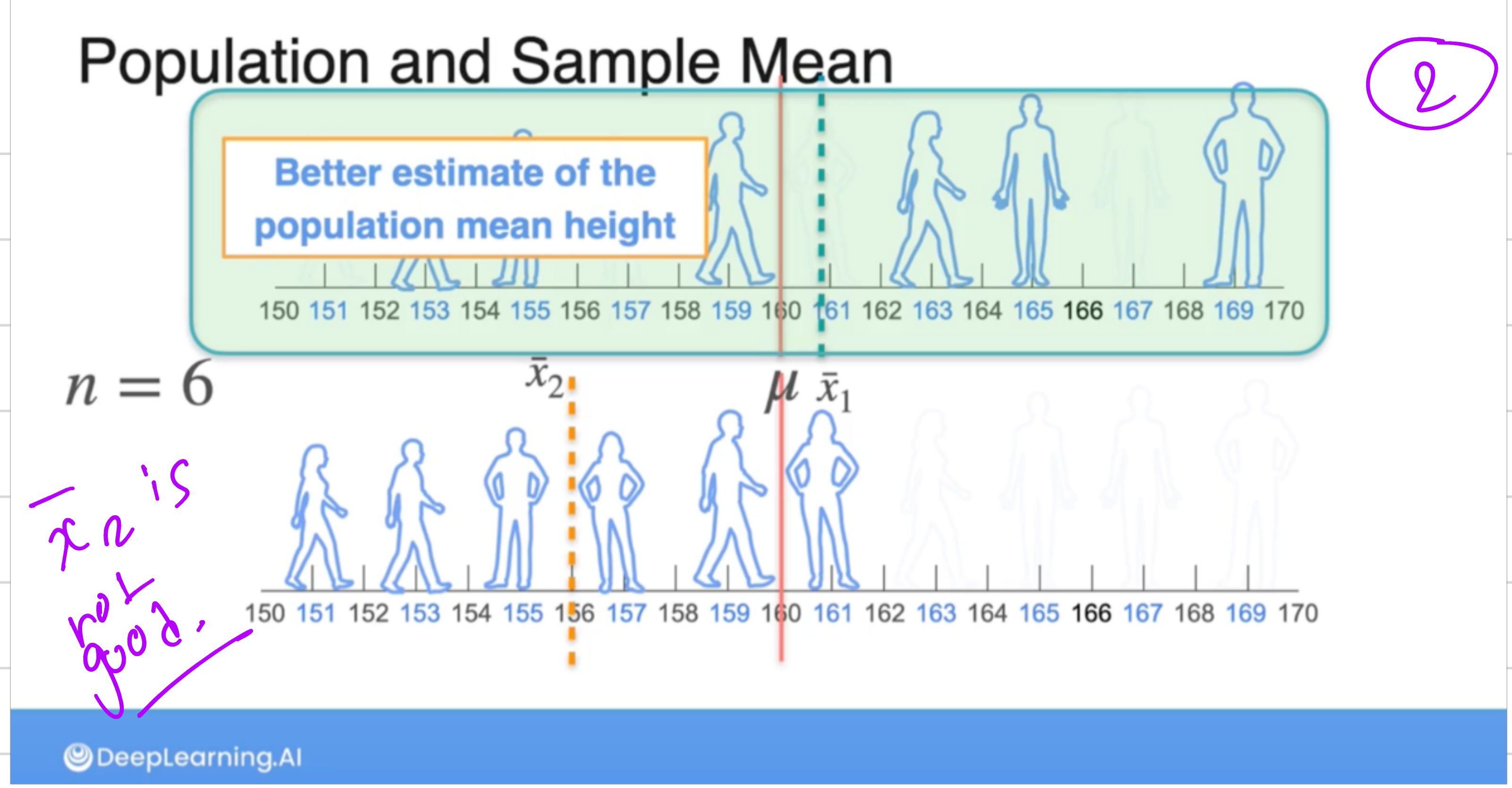
Correct

A random sample is more likely to represent the population because it avoids bias and ensures that every individual in the population has an equal chance of being selected. Therefore, the mean calculated from a random sample is more likely to be closer to the population mean.



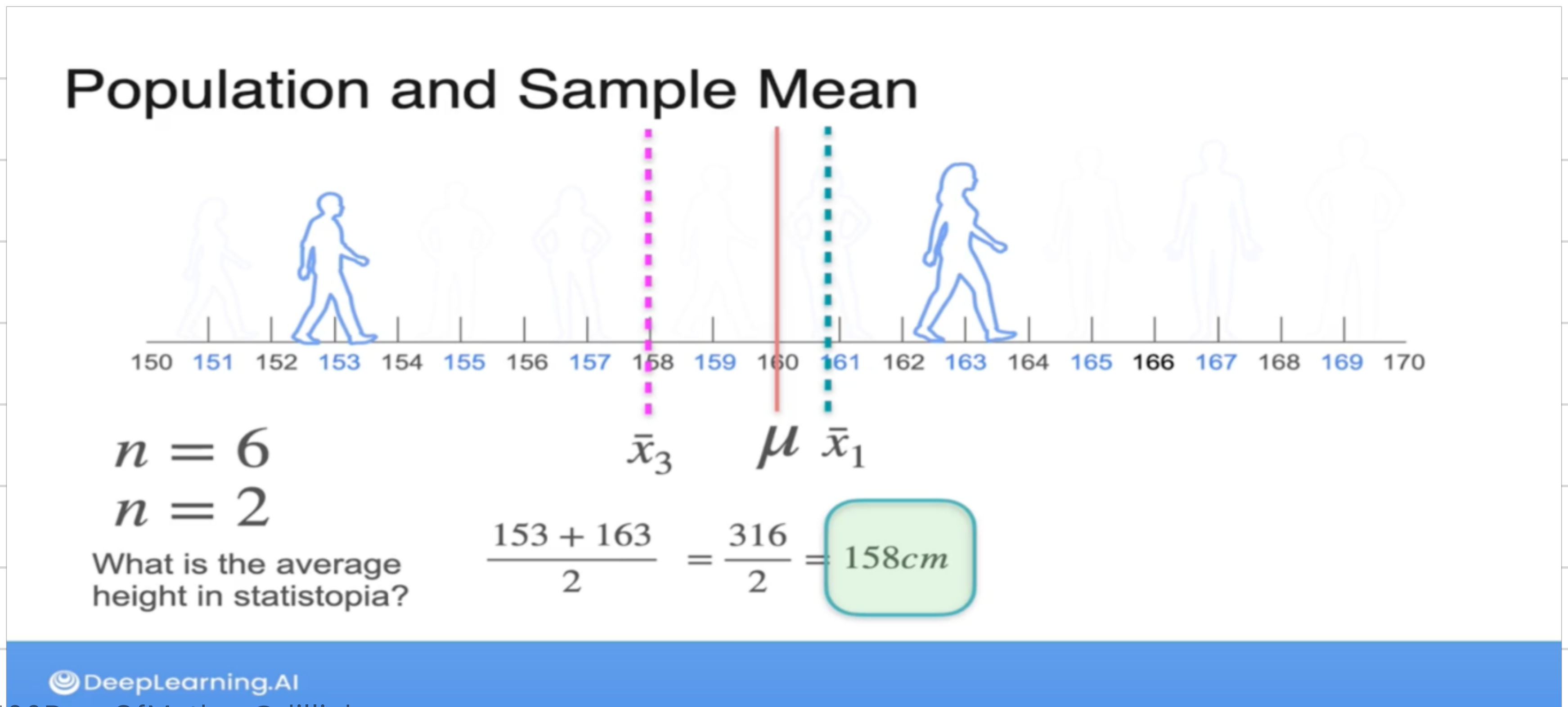






When comparing sample means, which statement is generally true when comparing a small sample to a large sample?

Larger samples tend to produce more reliable estimates of the population parameters because they reduce the impact of sampling variability. Therefore, the mean of a large sample is more likely to be closer to the population mean.



See, the shifting of center.

Proportion

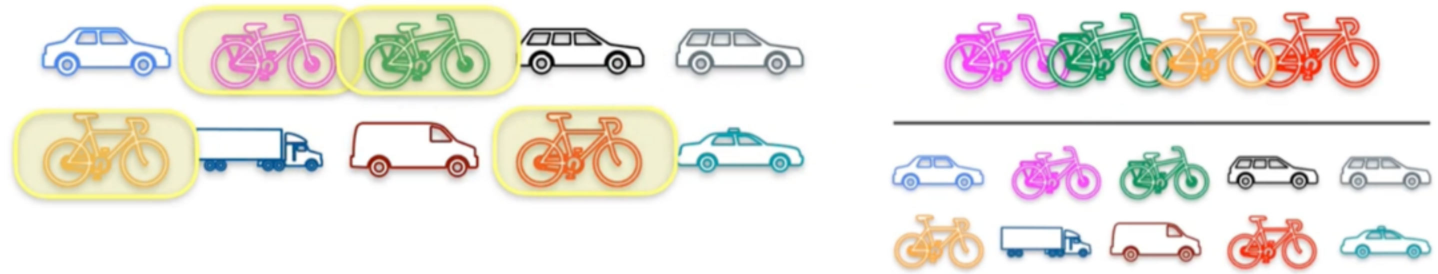
Population size: 10

What proportion of people own a bicycle?

DeepLearning.AI

Proportion

Population size: 10



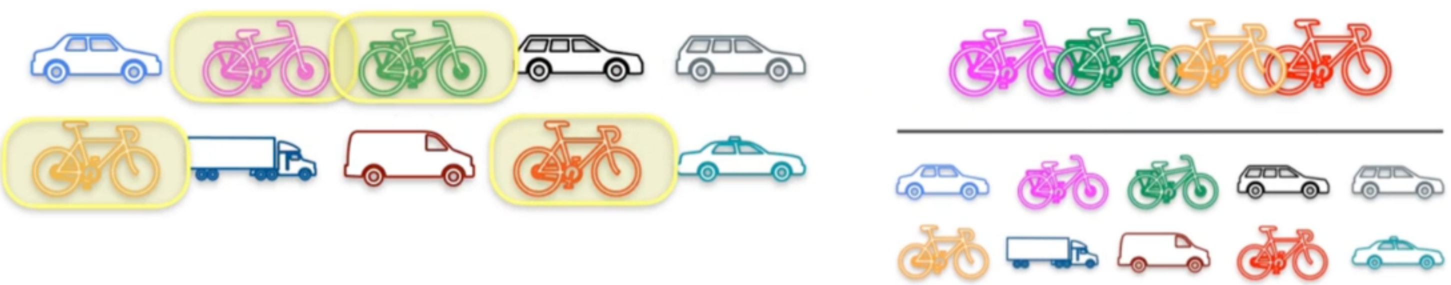
What proportion of people own a bicycle?

population proportion $= \frac{4}{10} = 0.4 = 40\,\%$

DeepLearning.AI

Proportion

Population size: 10



What proportion of people own a bicycle?

population proportion $= \frac{4}{10} = 0.4 = 40\,\%$

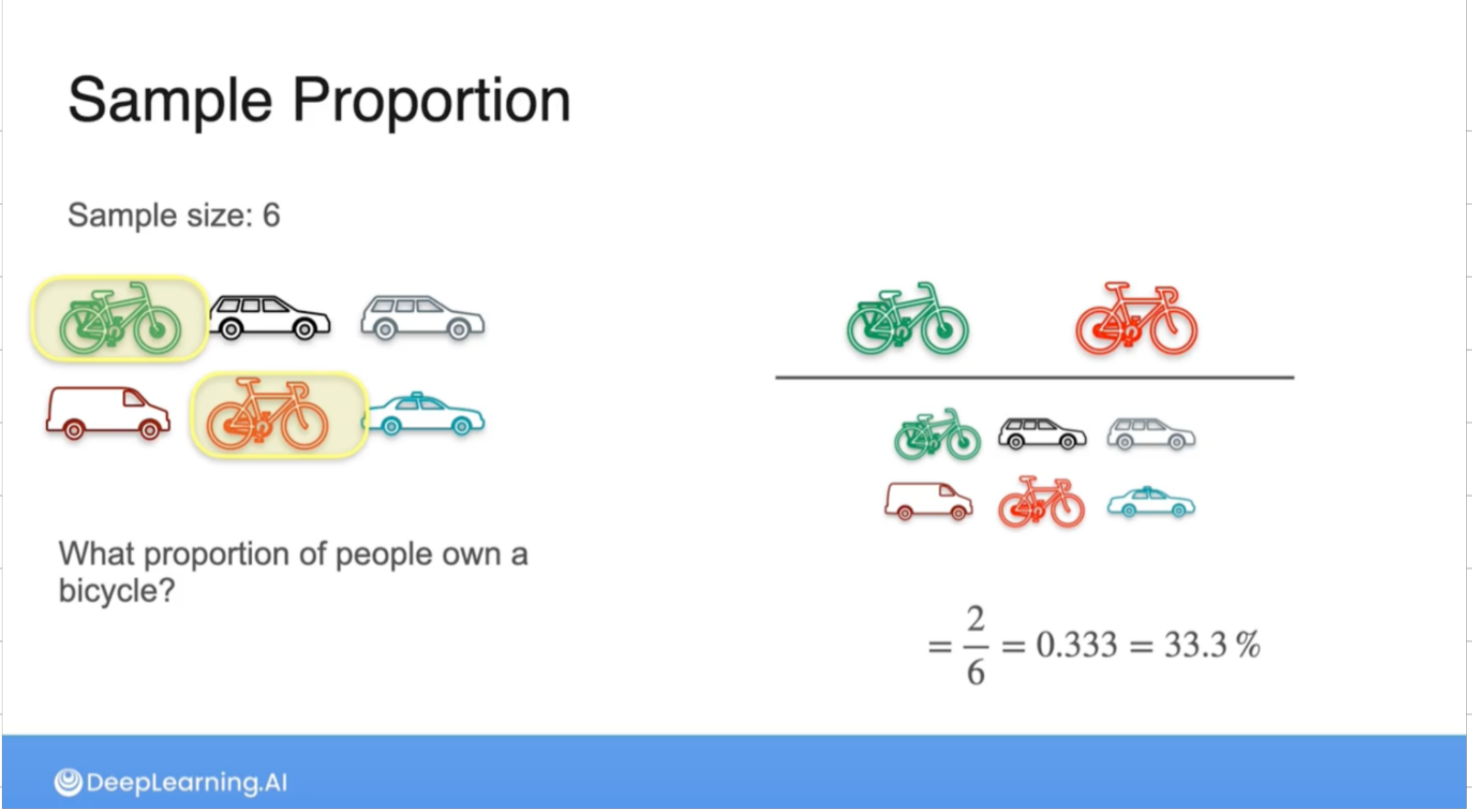
DeepLearning.AI

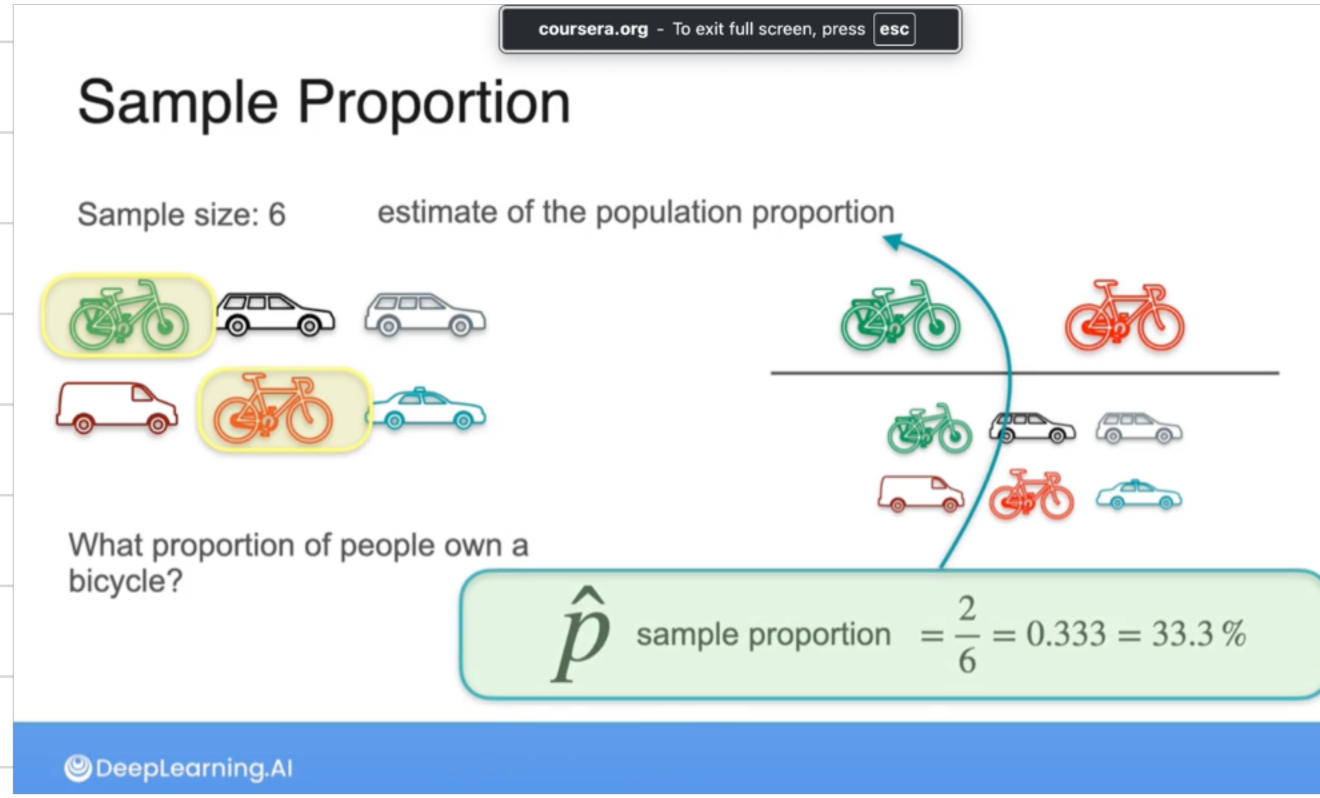
Proportion

population proportion

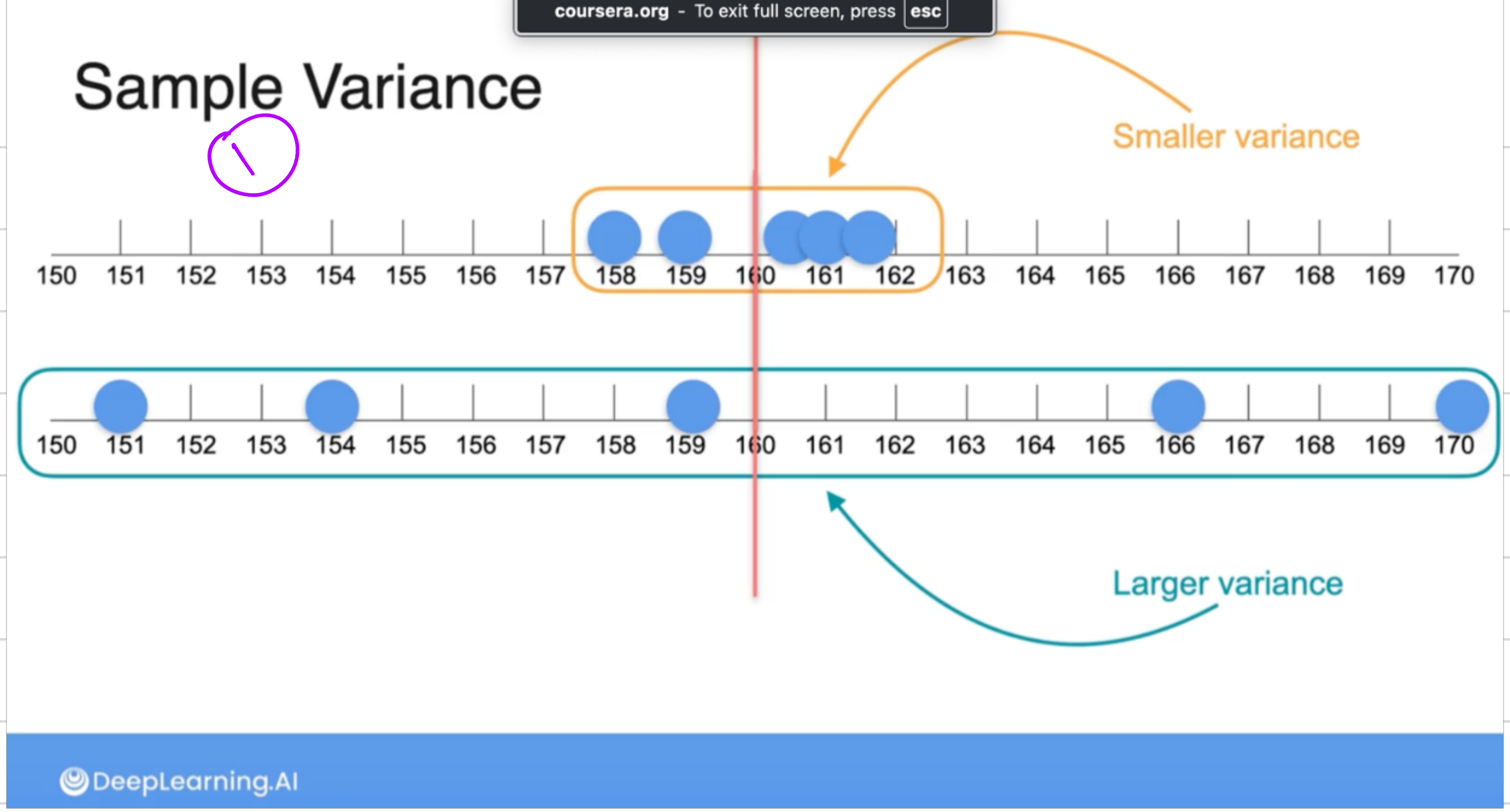
$$p = \frac{\text{number of items with a given characteristic } (x)}{\text{population } (N)}$$

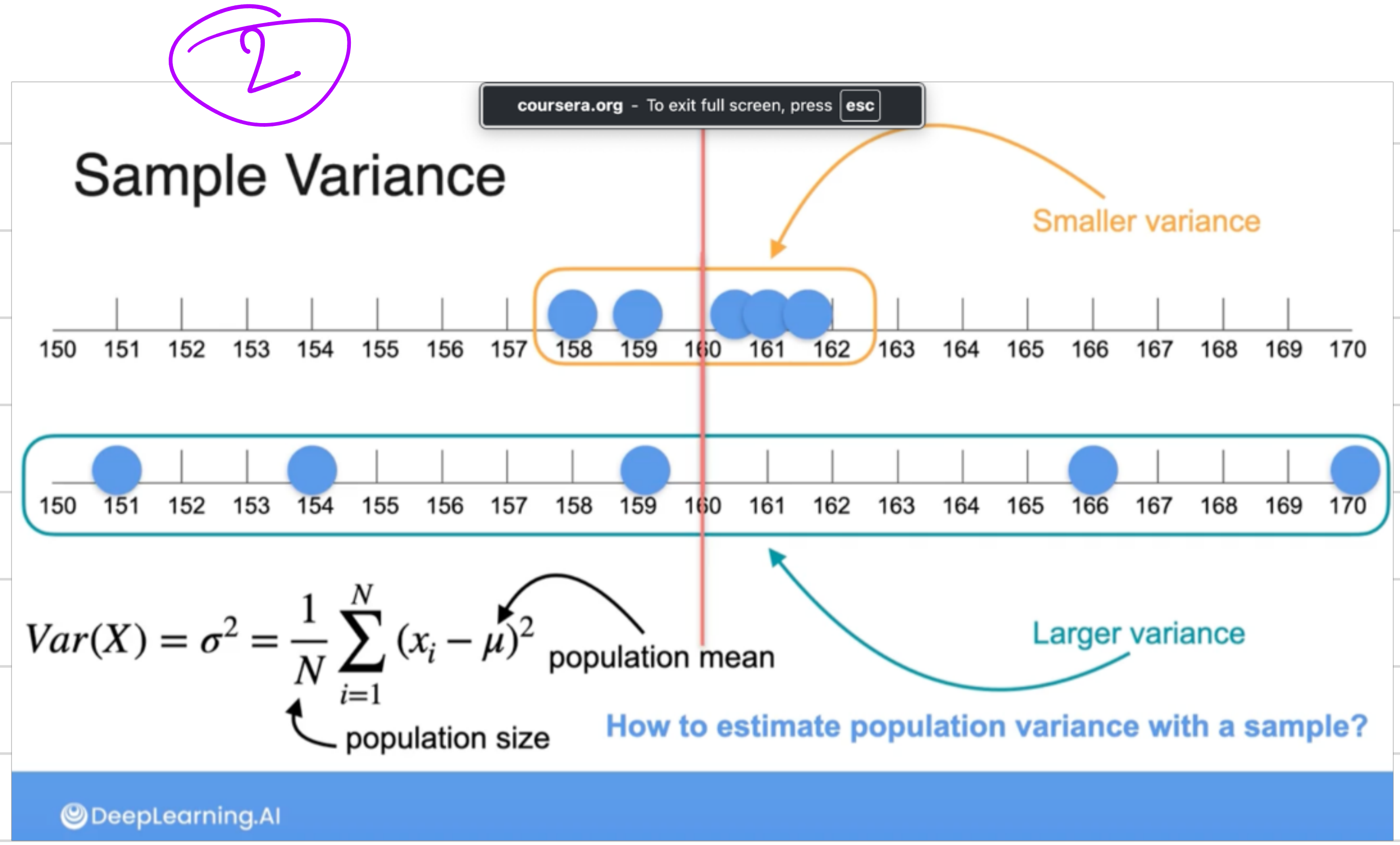
DeepLearning.AI

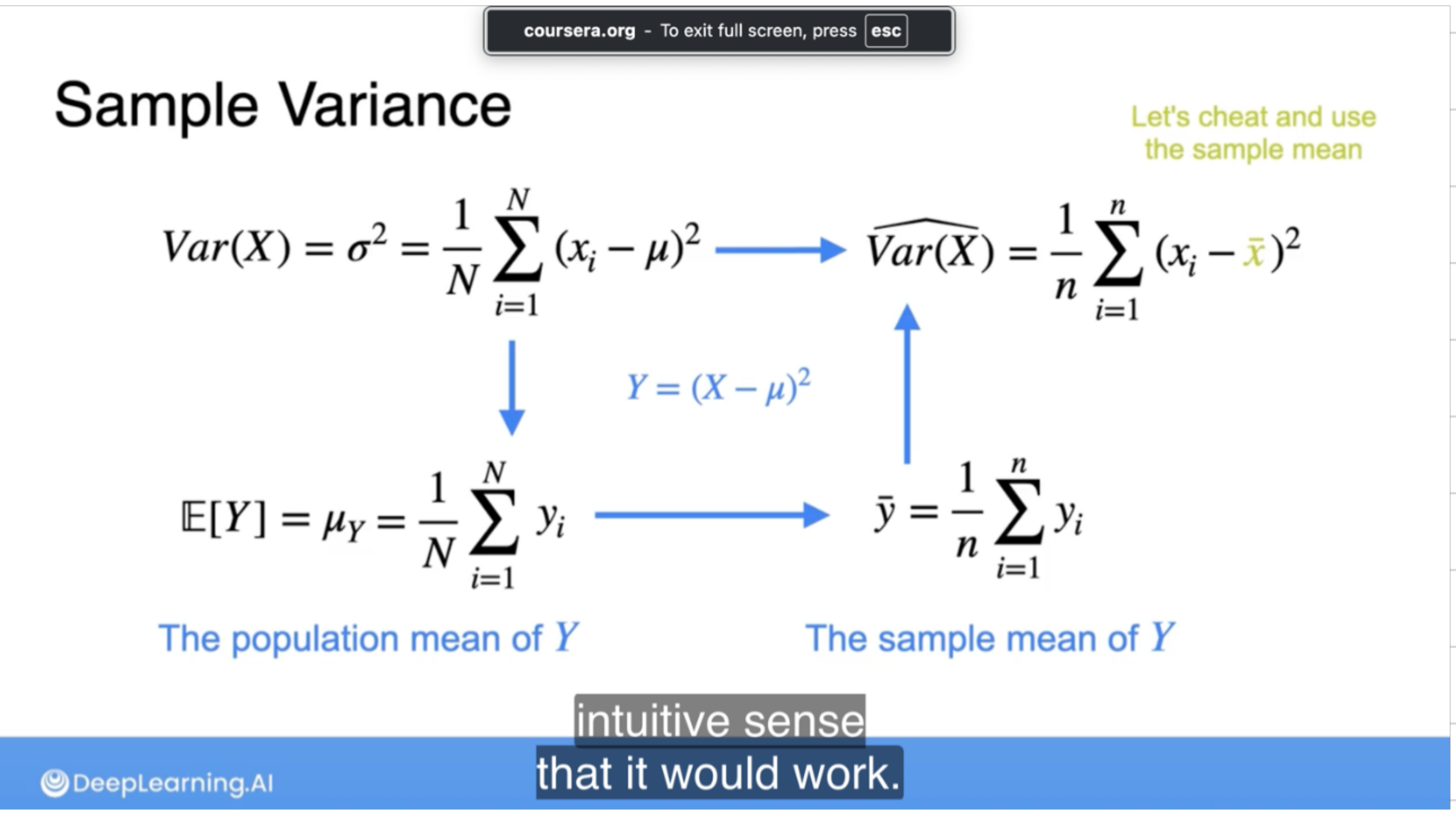


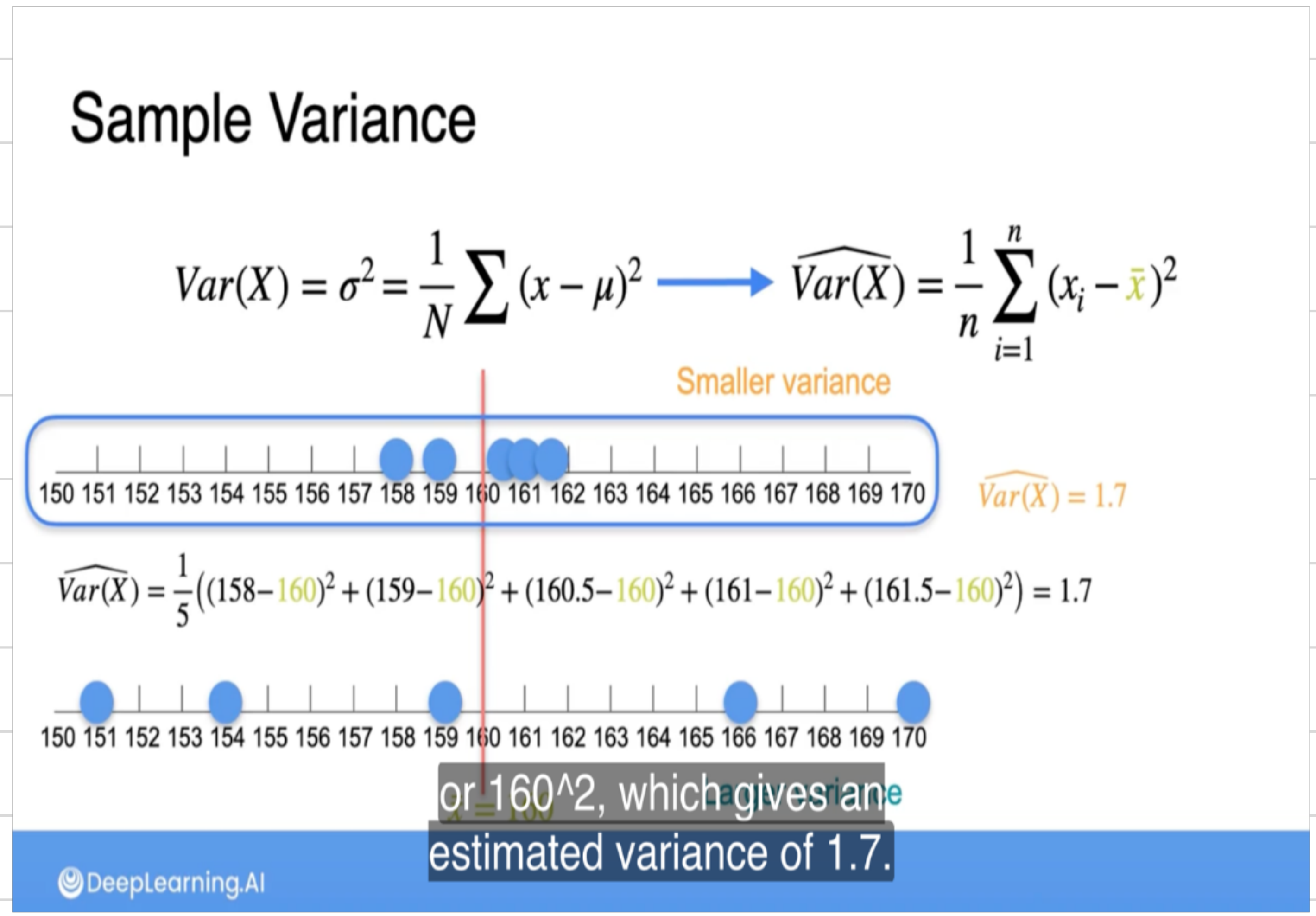


So the Idea is take population and randomly select the sample put some condition and find $\hat{p}$ sample proportion (very important for ML).

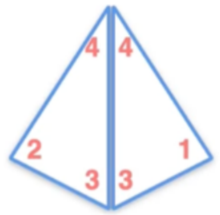

coursera.org - To exit full screen, press esc
Sample Variance
1
Smaller variance
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
Larger variance
DeepLearning.AI


2
coursera.org - To exit full screen, press esc
Sample Variance
Smaller variance
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
$Var(X) = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$
population mean
population size
Larger variance
How to estimate population variance with a sample?
DeepLearning.AI


coursera.org - To exit full screen, press esc
Sample Variance
Let's cheat and use the sample mean
$Var(X) = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \longrightarrow \widehat{Var(X)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
$Y = (X - \mu)^2$
$\mathbb{E}[Y] = \mu_Y = \frac{1}{N}\sum_{i=1}^{N} y_i \longrightarrow \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
The population mean of Y
The sample mean of Y
intuitive sense
that it would work.
DeepLearning.AI


Sample Variance
$Var(X) = \sigma^2 = \frac{1}{N}\sum(x - \mu)^2 \longrightarrow \widehat{Var(X)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Smaller variance
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
$\widehat{Var(X)} = 1.7$
$\widehat{Var(X)} = \frac{1}{5}\left((158-160)^2 + (159-160)^2 + (160.5-160)^2 + (161-160)^2 + (161.5-160)^2\right) = 1.7$
150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
or 160^2, which gives an
estimated variance of 1.7.
DeepLearning.AI

Law of Large Numbers



1 2 3 4

$\mu = 2.5$

Experiment:

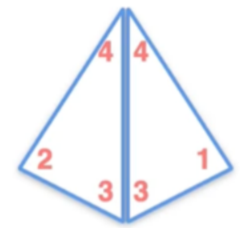
Toss the 4-sided dice twice and record the average of your outcomes

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

	1	2	3	4
1	1	1.5	2	2.5
2	1.5	2	2.5	3
3	2	2.5	3	3.5
4	2.5	3	3.5	4

DeepLearning.AI

Law of Large Numbers



1 2 3 4

$\mu = 2.5$

Experiment:

Toss the 4-sided dice twice and record the average of your outcomes

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

	1	2	3	4
1	1	1.5	2	2.5
2	1.5	2	2.5	3
3	2	2.5	3	3.5
4	2.5	3	3.5	4

	1	2	3	4
1	μ			
2	**2.5**			
3				
4				

DeepLearning.AI

Law of Large Numbers

1 2 3 4

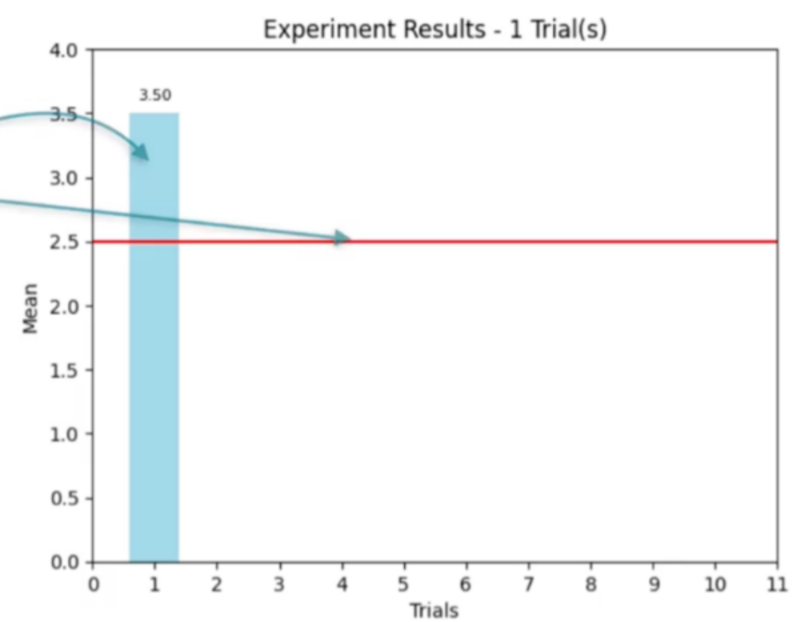
$\mu = 2.5$

1 trial

4,3

$\bar{x}_1$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



DeepLearning.AI

Law of Large Numbers

1 2 3 4

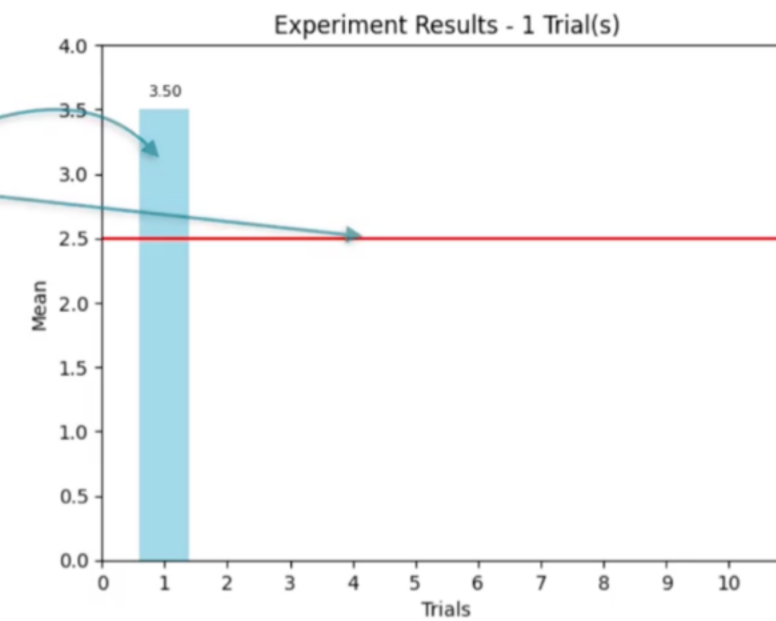
$\mu = 2.5$

1 trial

4,3

$\bar{x}_1$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4



DeepLearning.AI

Law of Large Numbers

1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

1 trial

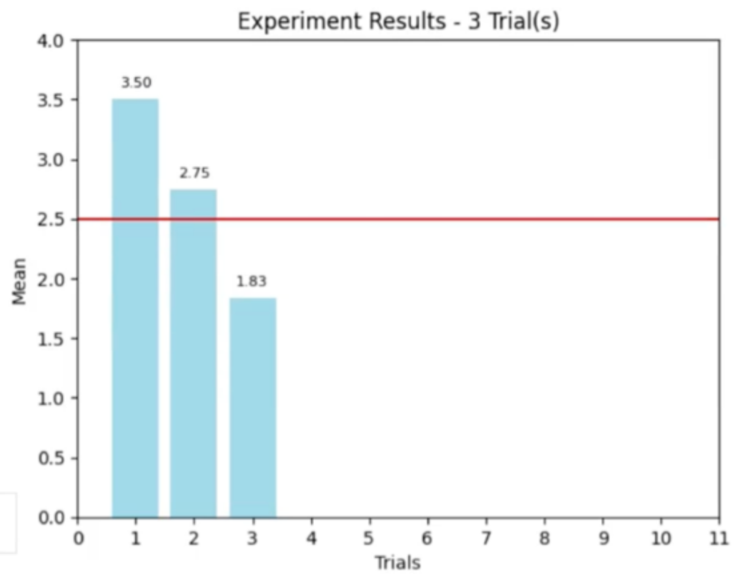
4,3

2 trials

3,4
1,3

3 trials

3,1 | 1,4 | 1,1



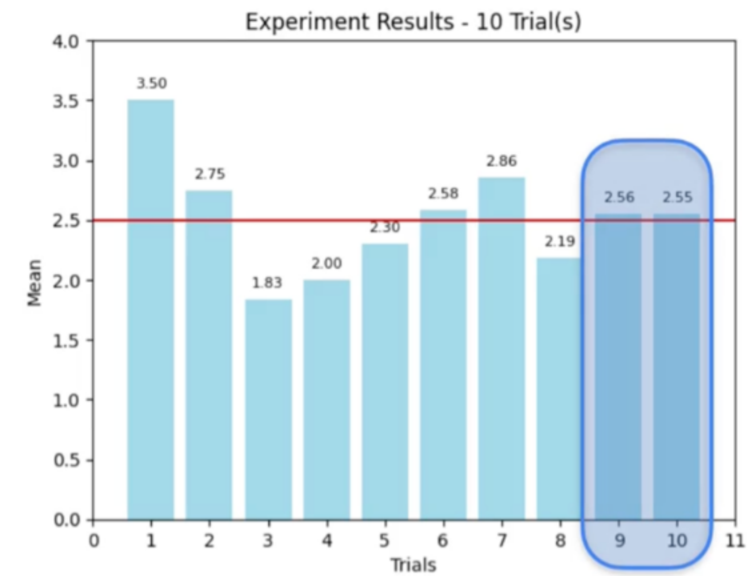
DeepLearning.AI

Law of Large Numbers

1 2 3 4

$\mu = 2.5$

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

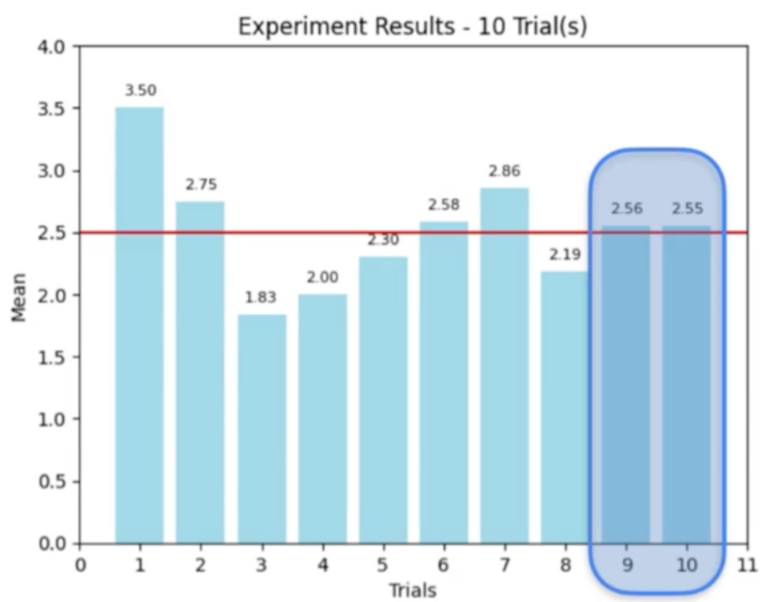


DeepLearning.AI

Law of Large Numbers

As the sample size increases, the average of the sample will tend to get closer to the average of the entire population.

Law of Large Numbers



DeepLearning.AI

Law of Large Numbers

Law of Large Numbers

as $n \to \infty$

$$\frac{X_1 + X_2 + X_3 + \ \ldots + \ X_n}{n} \longrightarrow \mathbb{E}[X] = \mu_X$$

n : number of samples

X_i : is the i-th random sample from the population.

Each X_i are independent and identically distributed (i.i.d.)

UNDER CERTAIN CONDITIONS

as $n \to \infty$ $\quad \frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X] = \mu_X$

DeepLearning.AI

Law of Large Numbers

UNDER CERTAIN CONDITIONS

- Sample is randomly drawn.
- Sample size must be sufficiently large.
- Independent observations.

DeepLearning.AI

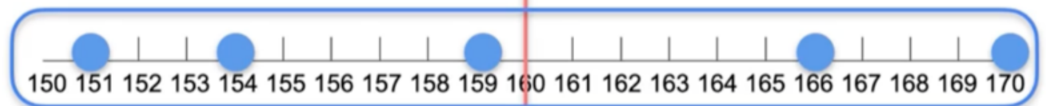
Sample Variance

$$Var(X) = \sigma^2 = \frac{1}{N}\sum (x-\mu)^2 \longrightarrow \widehat{Var(X)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Smaller variance

$\widehat{Var(X)} = 1.7$

$$\widehat{Var(X)} = \frac{1}{5}\left((151-160)^2 + (154-160)^2 + (159-160)^2 + (166-160)^2 + (170-160)^2\right) = 50.8$$

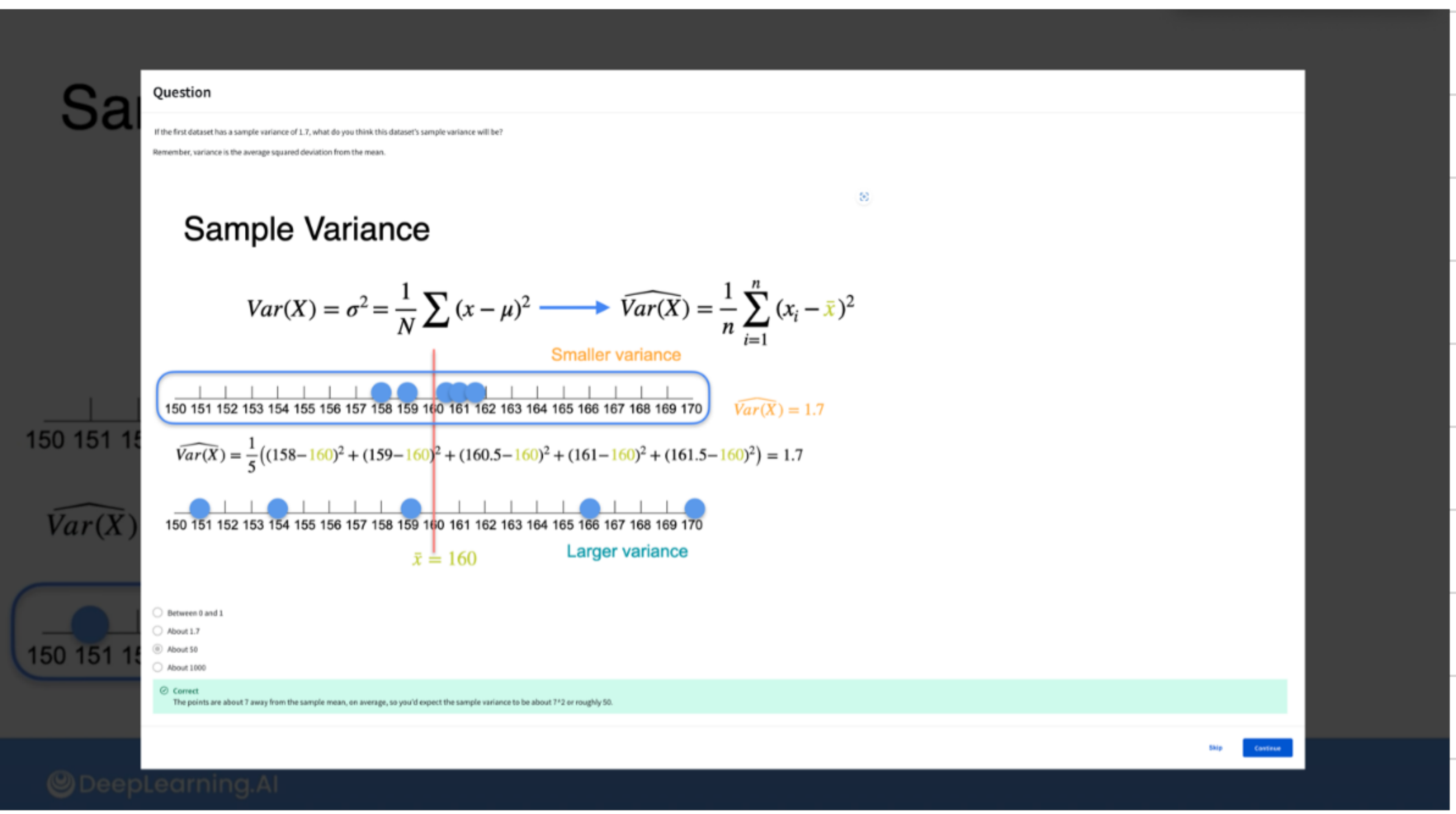


$\widehat{Var(X)} = 50.8$

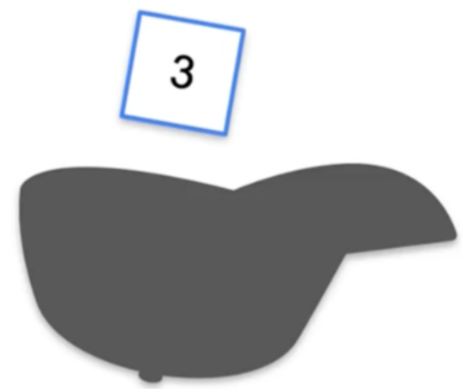
$\bar{x} = 160$

Larger variance

DeepLearning.AI



Variance Estimation



$$\mu = \frac{1+2+3}{3} = \frac{6}{3} = 2$$

+ 3/3 which gives you
6/3 which is simply 2.

DeepLearning.AI

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$x - 2$

x	$x-\mu$	$(x-\mu)^2$
1	-1	1
2	0	0
3	1	1

$$\frac{\sum (x-\mu)^2}{N} = \frac{2}{3}$$

σ^2
Population variance

the value you'll calculate for the population variance.

DeepLearning.AI

Variance Estimation

1	2	3

$\mu = 2$

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$\sigma^2 = \frac{2}{3}$$

$n=2$ Samples		$\bar{x}$	$\widehat{Var(X)} = \frac{\sum (x-\bar{x})^2}{n}$
1	1	1	0
1	2	1.5	0.25
1	3	2	1
2	1	1.5	0.25
2	2	2	0
2	3	2.5	0.25
3	1	2	1
3	2	2.5	0.25
3	3	3	0

estimated variance
$= 0.333$
$= \frac{1}{3}$

Clearly, there was an error here.

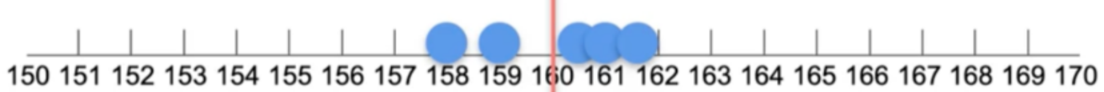
DeepLearning.AI

Sample Variance

This equation is "biased"
It underestimates the population variance

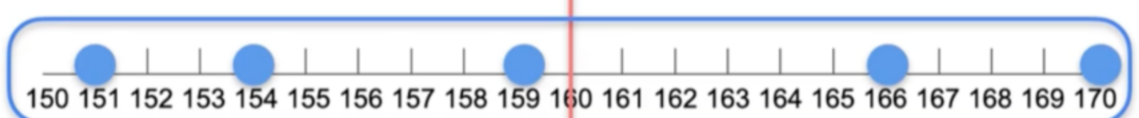
$$Var(X) = \sigma^2 = \frac{1}{N}\sum (x-\mu)^2 \longrightarrow \widehat{Var(X)} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Smaller variance



$\widehat{Var(X)} = 1.7$

$$\widehat{Var(X)} = \frac{1}{5}\left((151-160)^2 + (154-160)^2 + (159-160)^2 + (166-160)^2 + (170-160)^2\right) = 50.8$$



$\widehat{Var(X)} = 50.8$

$\bar{x} = 160$

Larger variance

will either over or

DeepLearning.AI

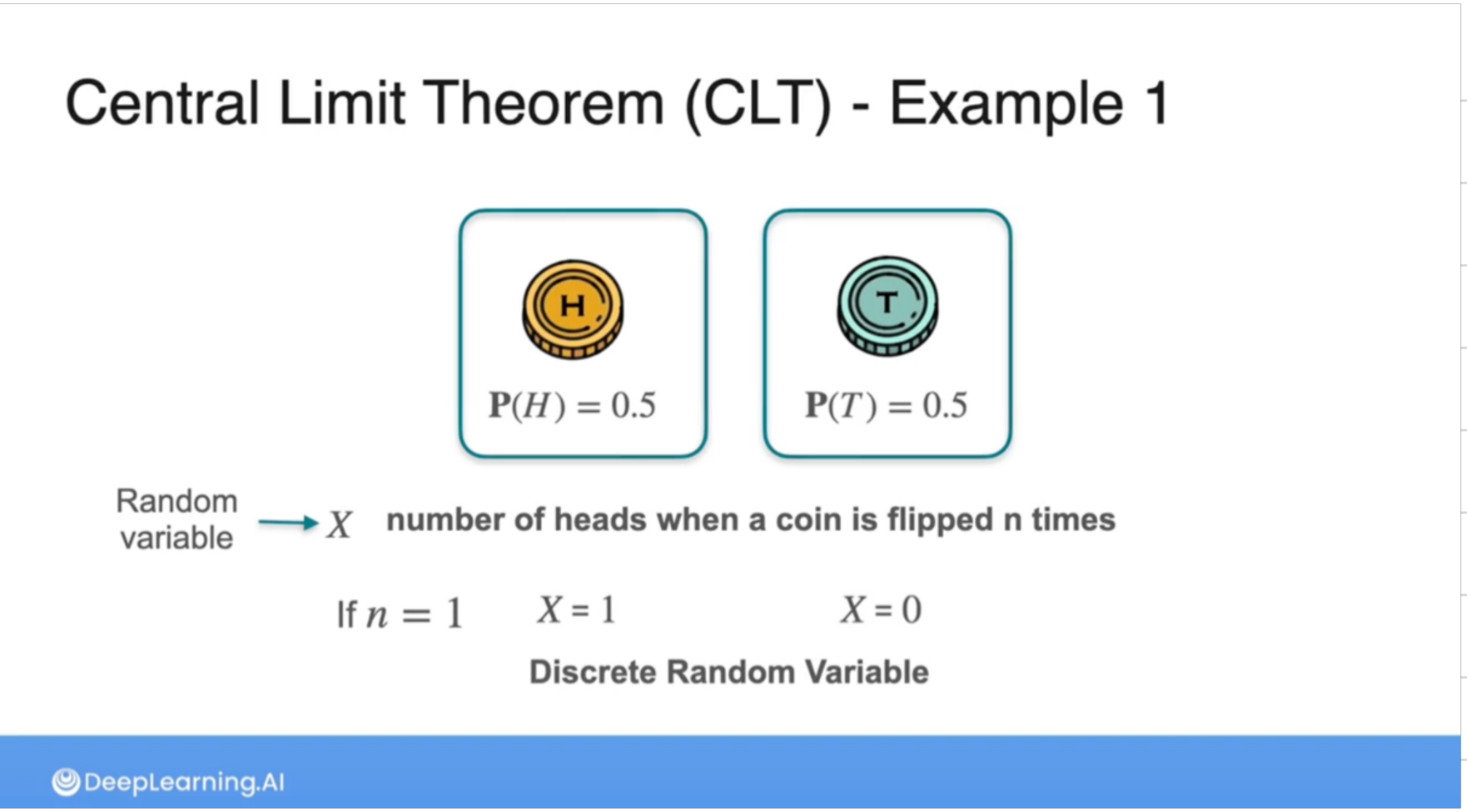
Variance Estimation

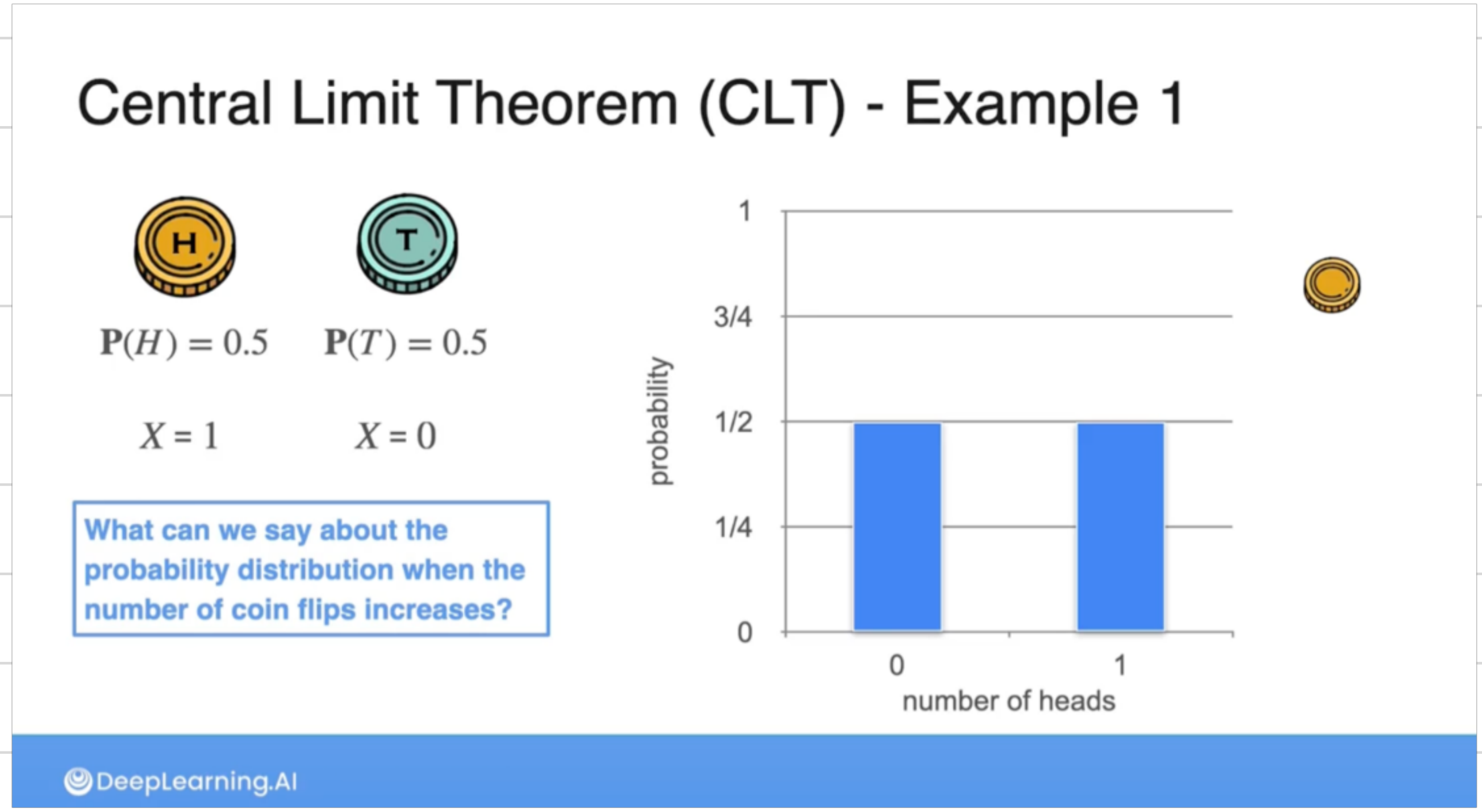
$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

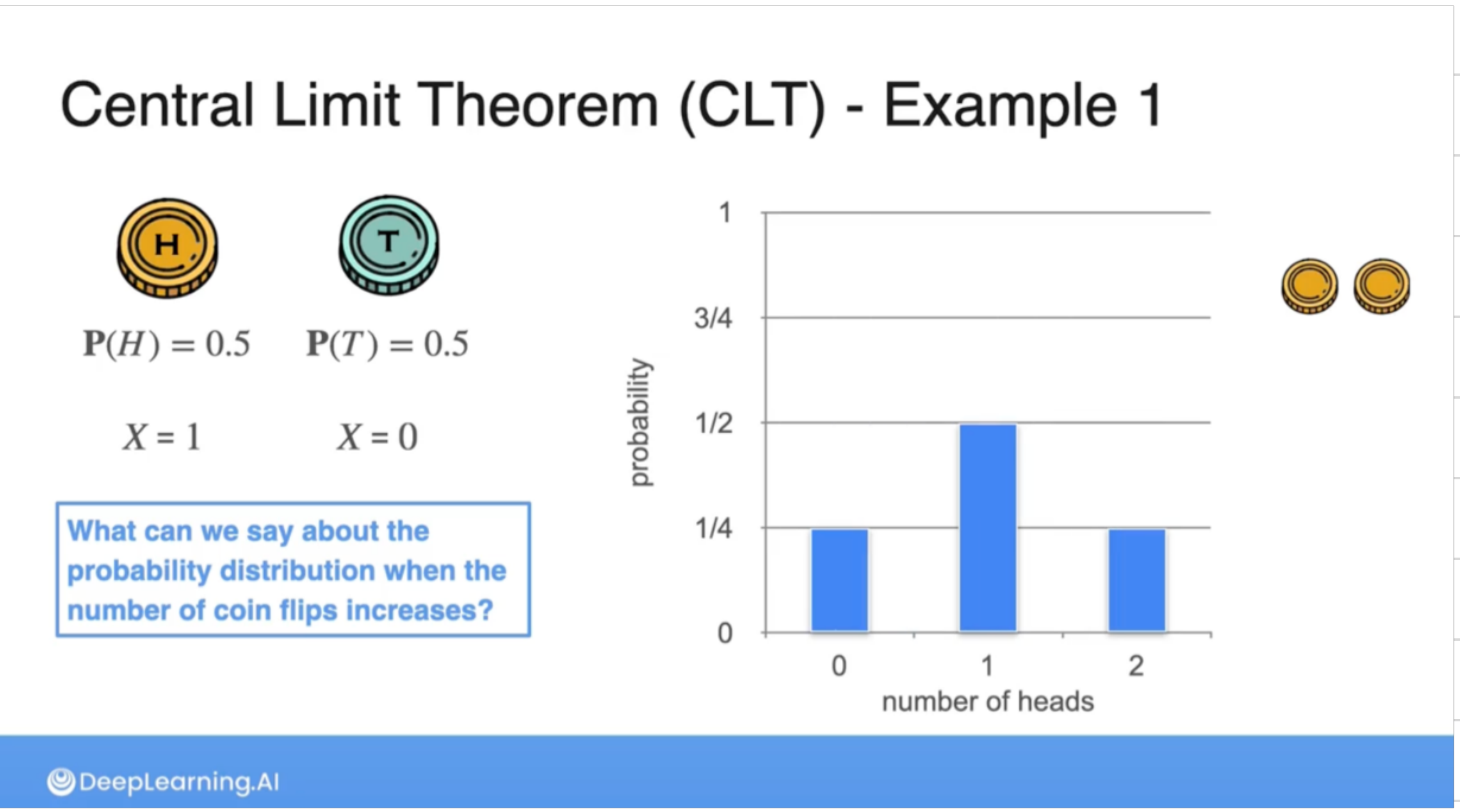
- n - 1 fixes bias when all you have is a sample
- As n gets big, the difference matters less
- If it matters, you may have too little data
- Some accepted statistical techniques use n

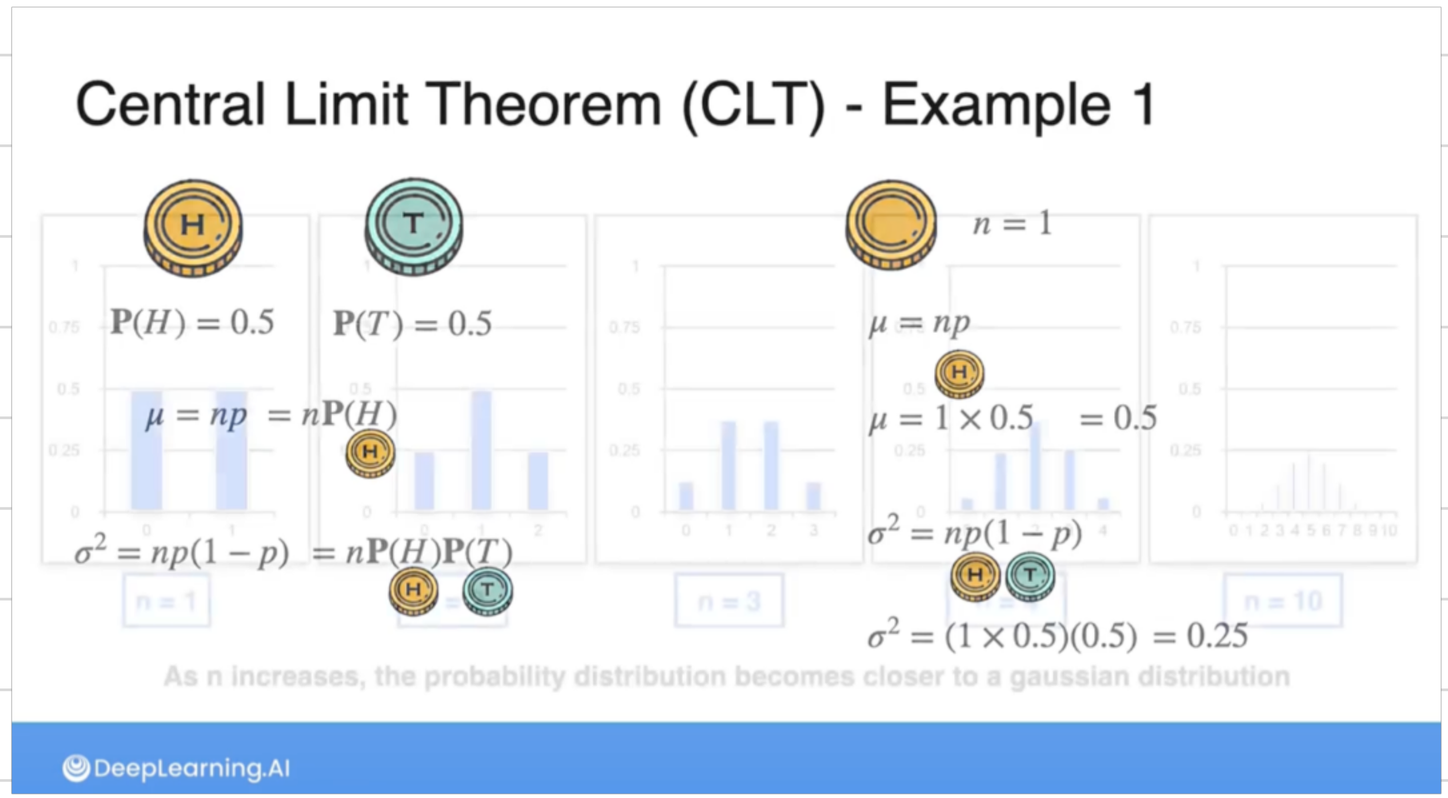
let's go back to the earlier examples

DeepLearning.AI


Central Limit Theorem (CLT) - Example 1
H
P(H) = 0.5
T
P(T) = 0.5
Random variable
X number of heads when a coin is flipped n times
If n = 1
X = 1
X = 0
Discrete Random Variable
DeepLearning.AI


Central Limit Theorem (CLT) - Example 1
H
T
P(H) = 0.5
P(T) = 0.5
X = 1
X = 0
What can we say about the probability distribution when the number of coin flips increases?
probability
1
3/4
1/2
1/4
0
0
1
number of heads
DeepLearning.AI


Central Limit Theorem (CLT) - Example 1
H
T
P(H) = 0.5
P(T) = 0.5
X = 1
X = 0
What can we say about the probability distribution when the number of coin flips increases?
probability
1
3/4
1/2
1/4
0
0
1
2
number of heads
DeepLearning.AI

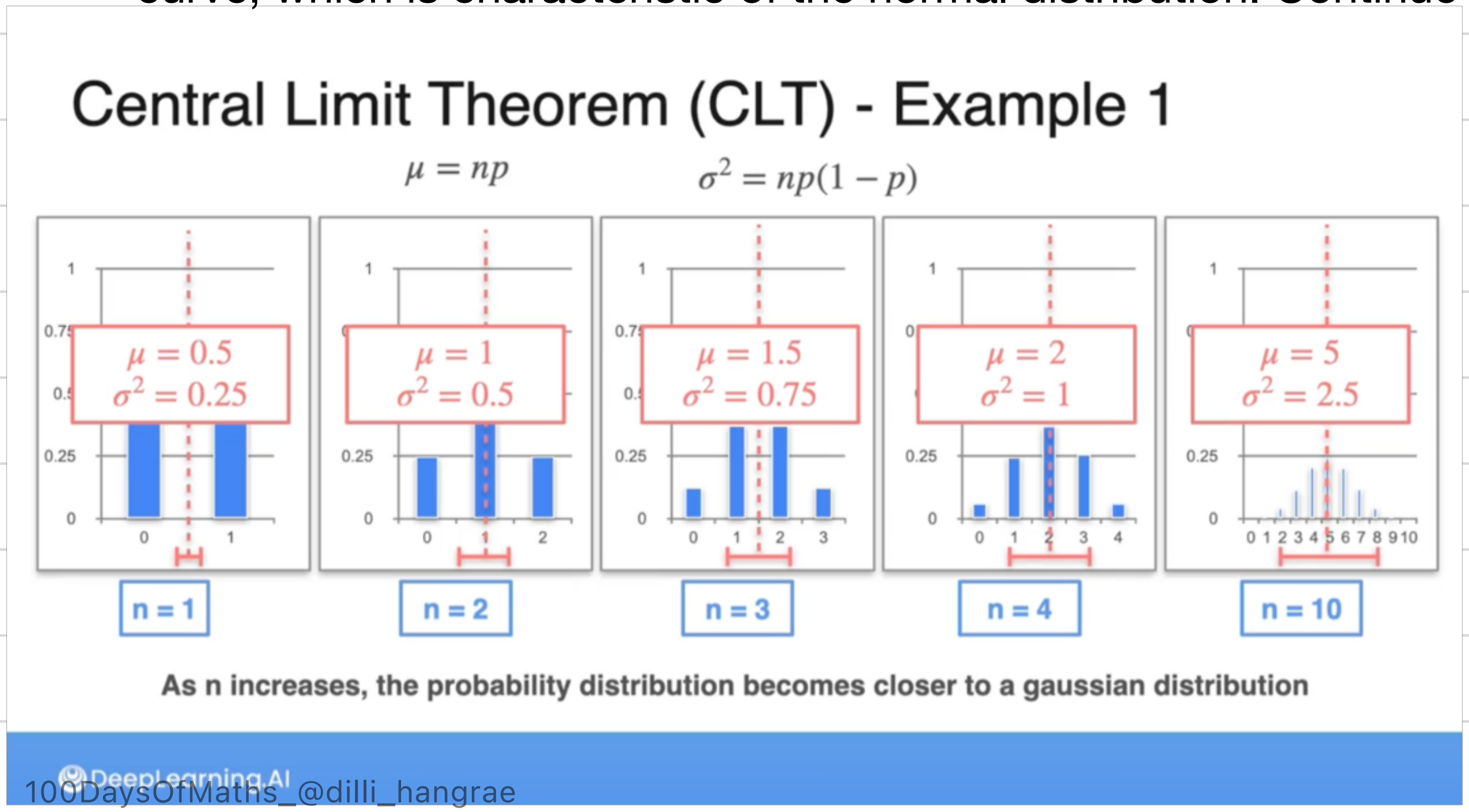

Central Limit Theorem (CLT) - Example 1
H
T
P(H) = 0.5
P(T) = 0.5
n = 1
μ = np = nP(H)
μ = np
μ = 1 × 0.5 = 0.5
σ² = np(1 − p) = nP(H)P(T)
σ² = np(1 − p)
σ² = (1 × 0.5)(0.5) = 0.25
n = 1
n = 3
n = 10
As n increases, the probability distribution becomes closer to a gaussian distribution
DeepLearning.AI

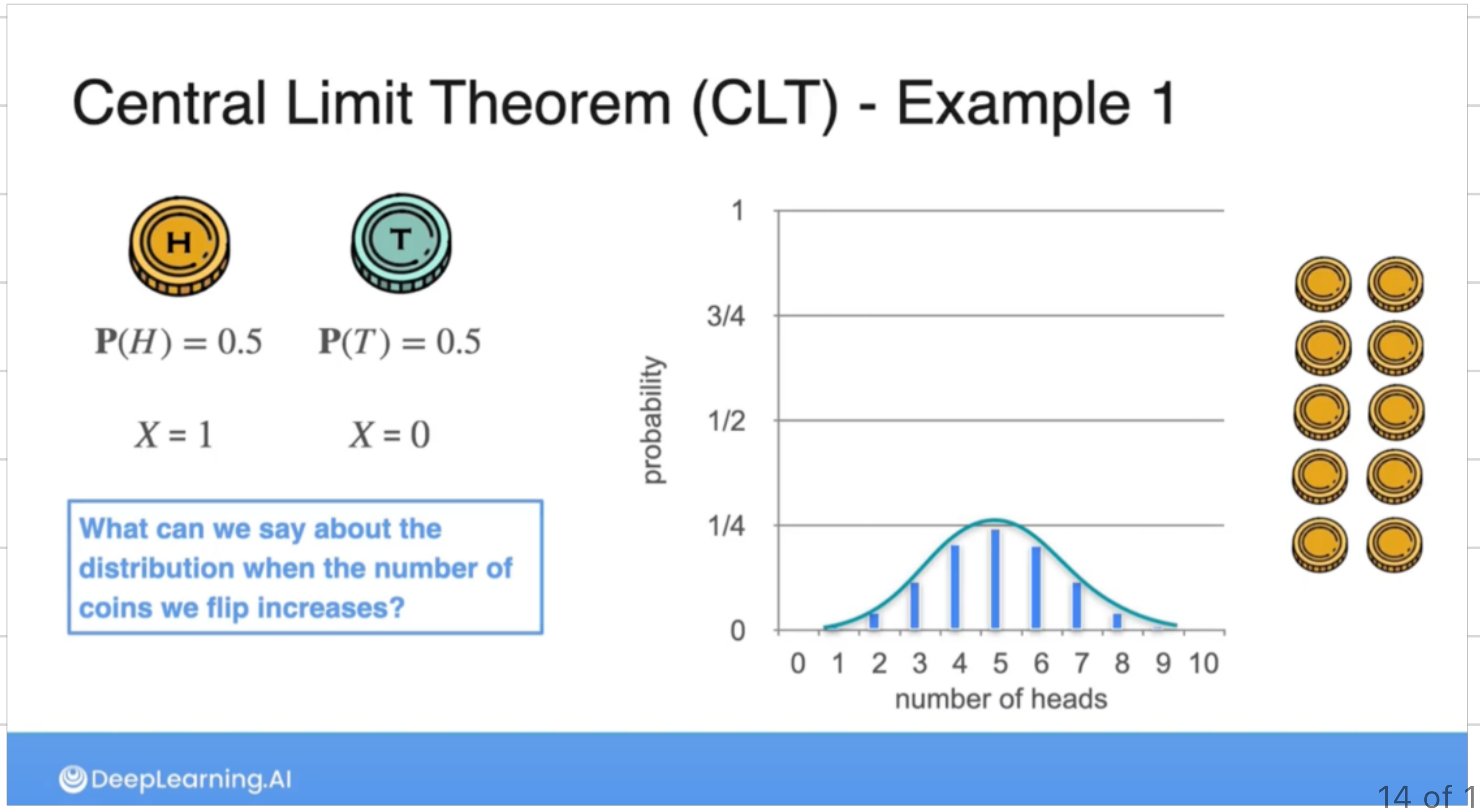
Remember, I told you before that the normal distribution appears in a lot of places. Here's one that perhaps you didn't expect. Take a distribution, any distribution. It can be as skewed as you want. Now, take a few samples, always the same number, and look at the average, and do this many times, and plot all these averages. Guess what you get? Yes, you get the normal distribution, no matter what distribution you started with in the first place. This is a fascinating result and is one of the pinnacles of statistics. It is called the central limit period.

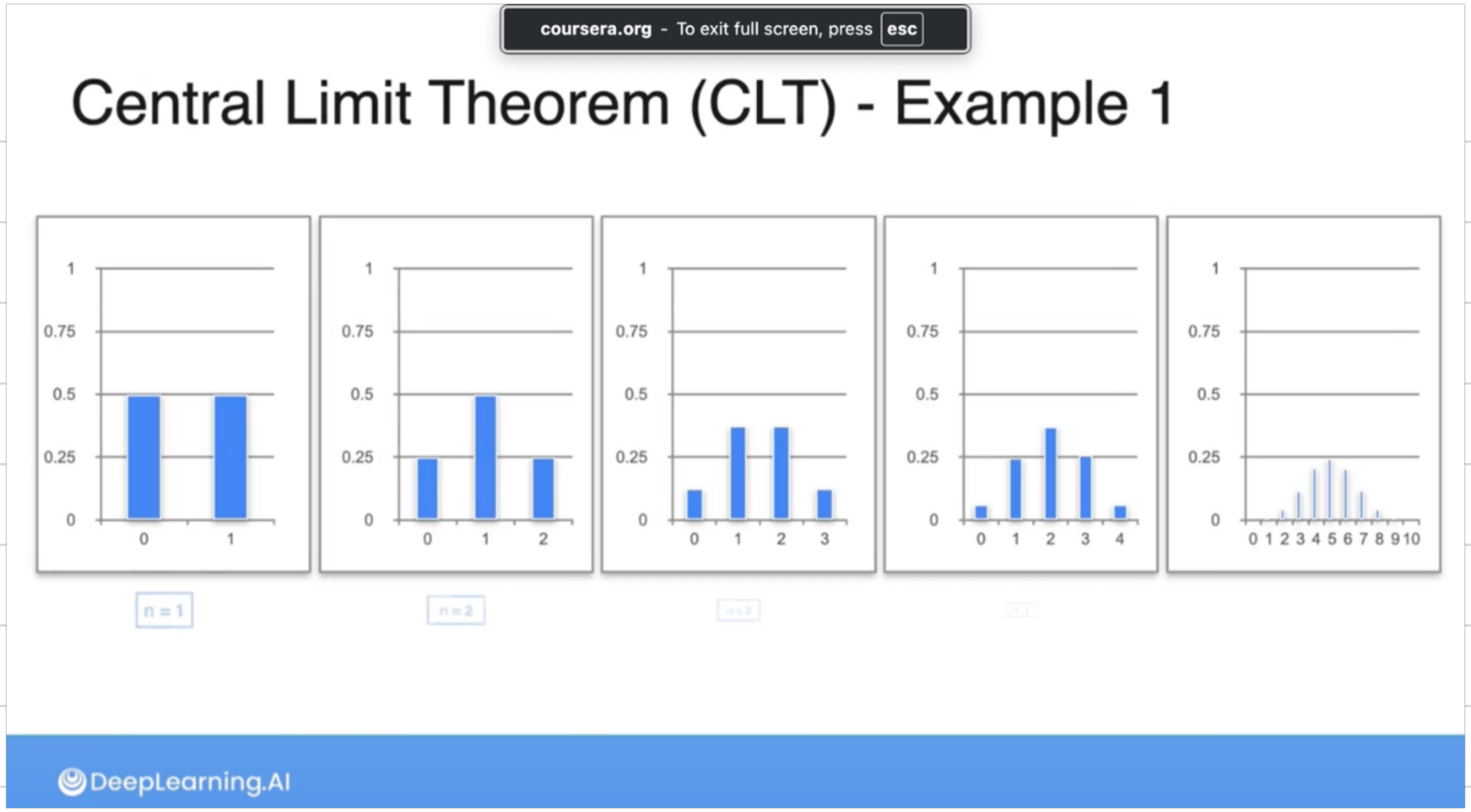
What can we say about the probability distribution of the number of heads flipped when the number of coin flips increases?

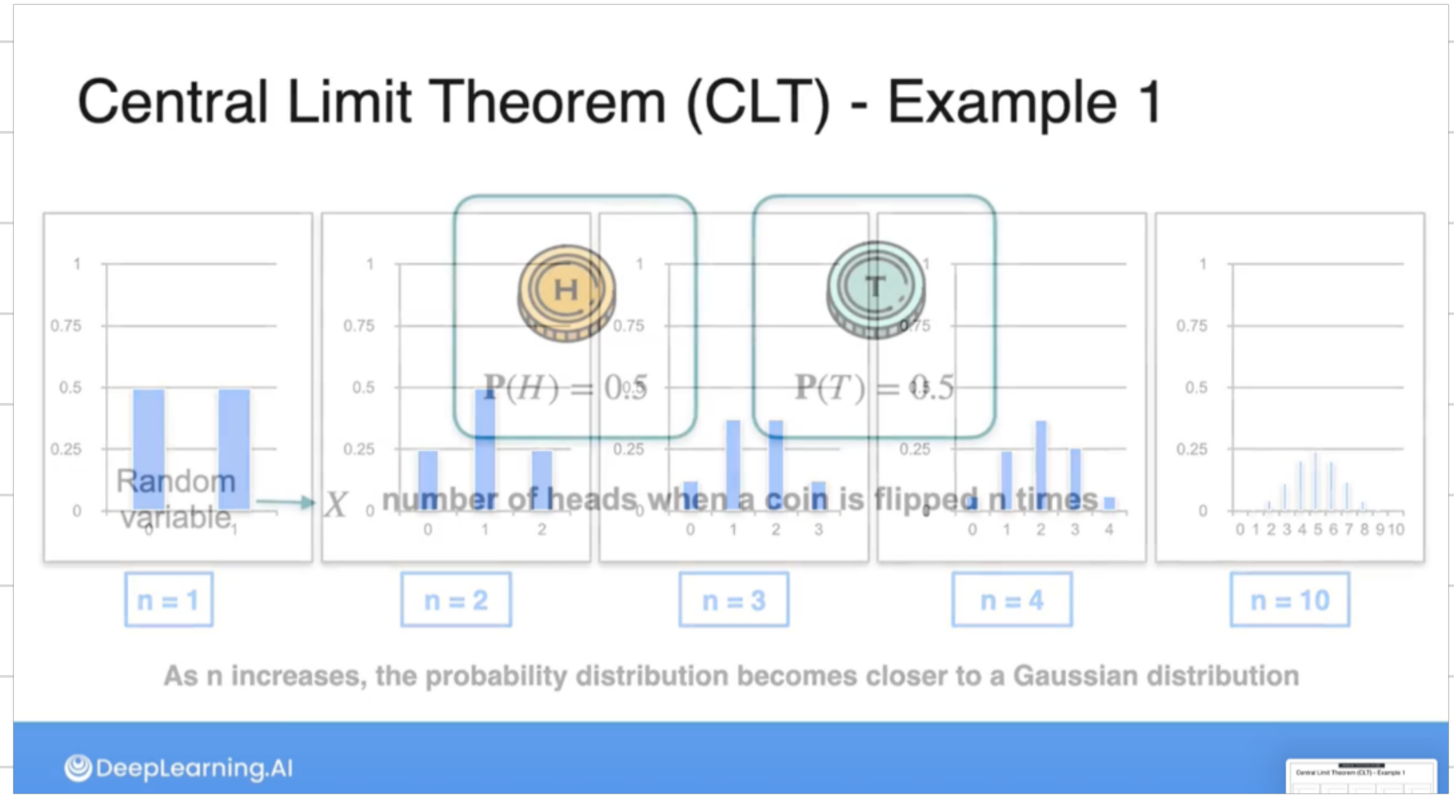
The probability distribution becomes more positively skewed.
The probability distribution becomes bell-shaped.
The probability distribution becomes more negatively skewed.
The probability distribution becomes more uniform.

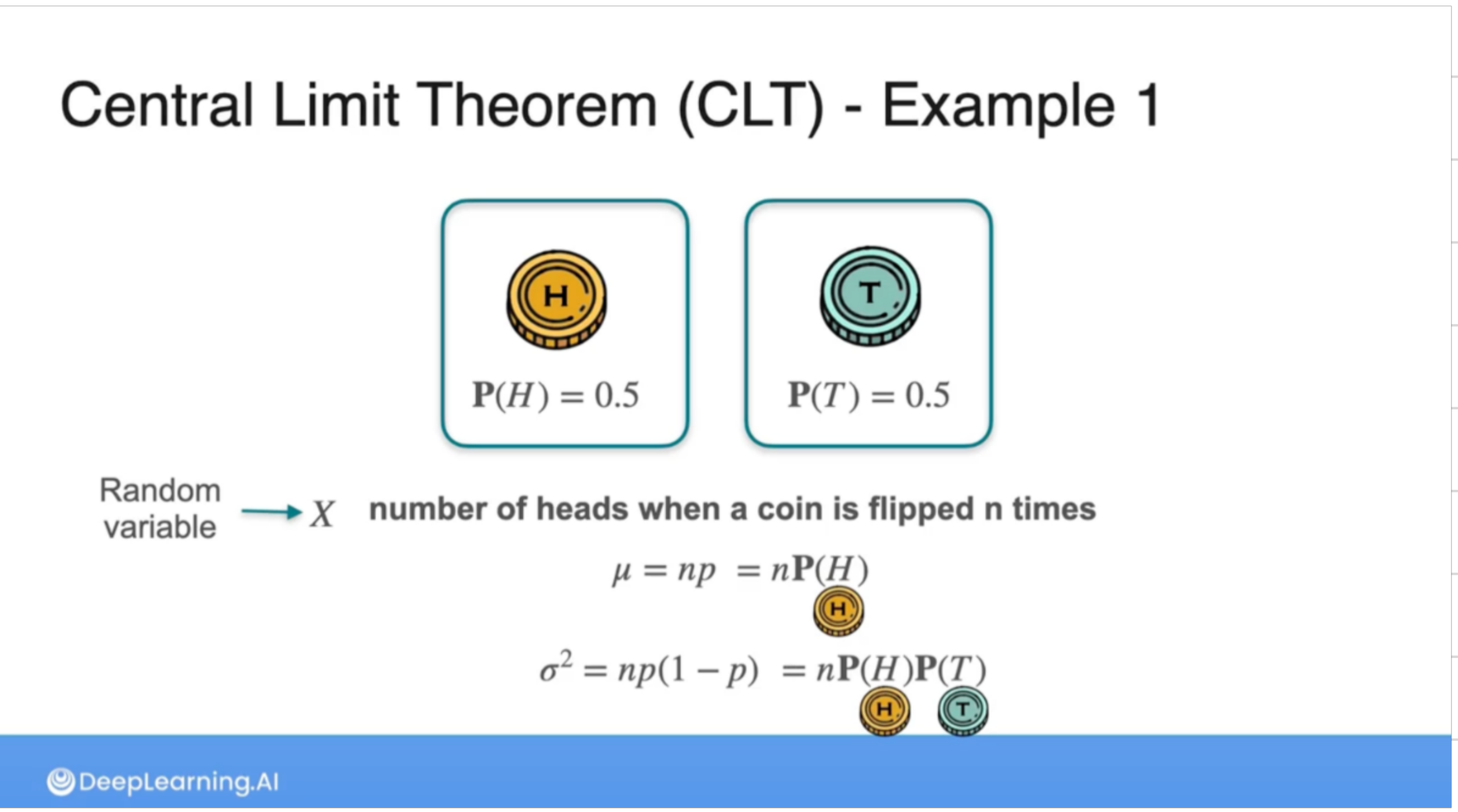
Correct
Great job! As the number of coin flips increases, the probability distribution of the number of heads flipped tends to follow a bell-shaped curve, which is characteristic of the normal distribution. Continue the video to hear the explanation.

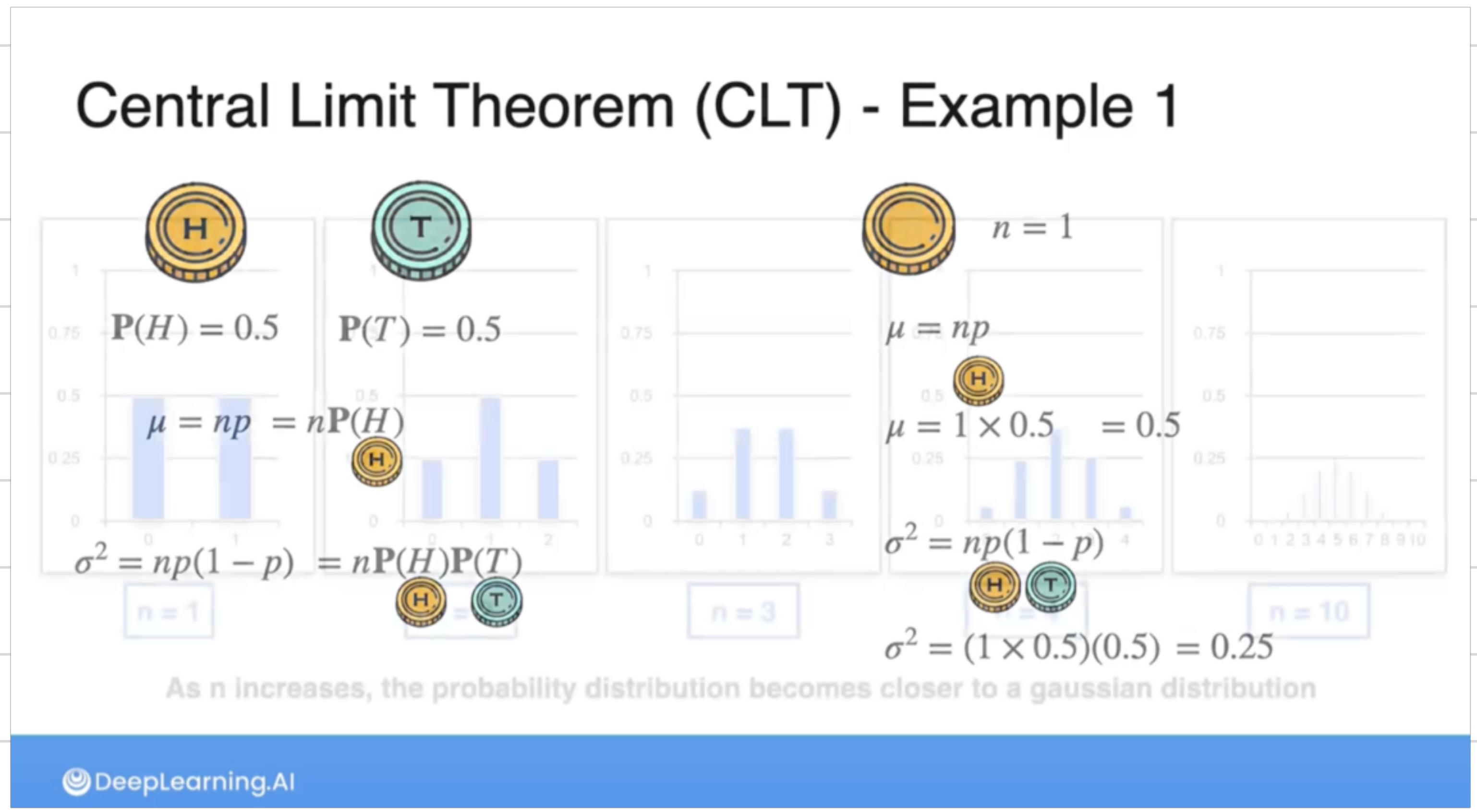





coursera.org - To exit full screen, press esc
Central Limit Theorem (CLT) - Example 1
n = 1
n = 2
DeepLearning.AI


Central Limit Theorem (CLT) - Example 1
$\mathbf{P}(H) = 0.5$
$\mathbf{P}(T) = 0.5$
Random variable → X number of heads when a coin is flipped n times
n = 1
n = 2
n = 3
n = 4
n = 10
As n increases, the probability distribution becomes closer to a Gaussian distribution
DeepLearning.AI


Central Limit Theorem (CLT) - Example 1
$\mathbf{P}(H) = 0.5$
$\mathbf{P}(T) = 0.5$
Random variable → X number of heads when a coin is flipped n times
$\mu = np = n\mathbf{P}(H)$
$\sigma^2 = np(1-p) = n\mathbf{P}(H)\mathbf{P}(T)$
DeepLearning.AI


Central Limit Theorem (CLT) - Example 1
$\mathbf{P}(H) = 0.5$
$\mathbf{P}(T) = 0.5$
$n = 1$
$\mu = np$
$\mu = np = n\mathbf{P}(H)$
$\mu = 1 \times 0.5 = 0.5$
$\sigma^2 = np(1-p) = n\mathbf{P}(H)\mathbf{P}(T)$
$\sigma^2 = np(1-p)$
$\sigma^2 = (1 \times 0.5)(0.5) = 0.25$
As n increases, the probability distribution becomes closer to a gaussian distribution
DeepLearning.AI

Central Limit Theorem (CLT) - Example 1

$\mu = np$ $\quad\quad$ $\sigma^2 = np(1-p)$

$\mu = 0.5$, $\sigma^2 = 0.25$	$\mu = 1$, $\sigma^2 = 0.5$	$\mu = 1.5$, $\sigma^2 = 0.75$	$\mu = 2$, $\sigma^2 = 1$	$\mu = 5$, $\sigma^2 = 2.5$

as n become sufficiently large we will get a normal distribution with parameters

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

DeepLearning.AI

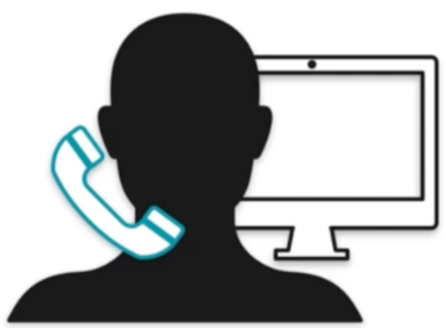
Uniform Distribution: Motivation

You're calling a tech support line.
They can answer any time between zero and 15 minutes and if they don't answer in this time, the line is disconnected.

$X =$ "Wait time for a called to be answered "

DeepLearning.AI

Uniform Distribution: Motivation



You're calling a tech support line.
They can answer any time between zero and 15 minutes and if they don't answer in this time, the line is disconnected.

$X =$ "Wait time for a called to be answered "

$$X \sim \mathcal{U}(0,15)$$

DeepLearning.AI

Central Limit Theorem (CLT) - Example 2

$n = 1$ $\quad$ $Y_1 = \frac{X_1}{1}$

$n = 2$ $\quad$ $Y_2 = \frac{X_1 + X_2}{2}$

$n = 3$ $\quad$ $Y_3 = \frac{X_1 + X_2 + X_3}{3}$

$\vdots$ $\quad$ $\vdots$

Record the average of all n experiments

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

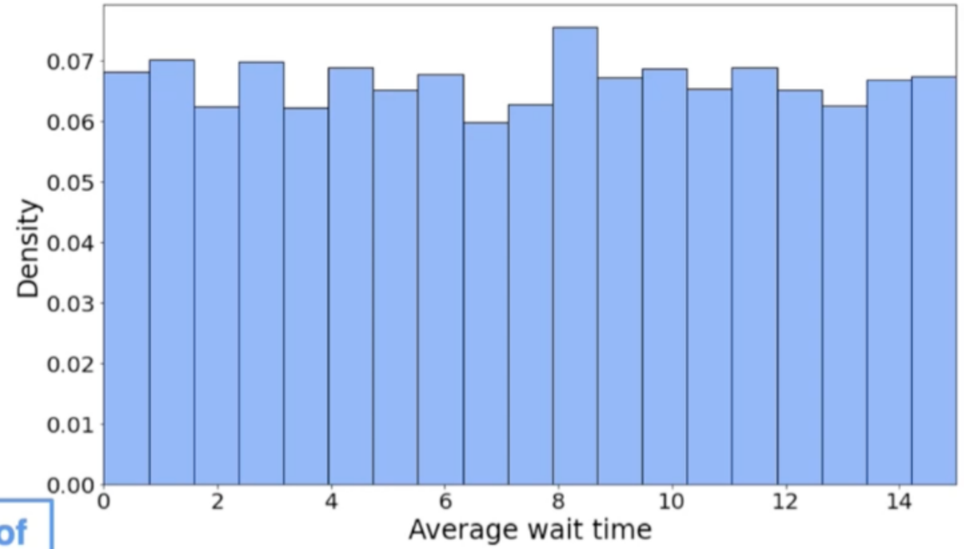
Can we say anything about the distribution of this average?

DeepLearning.AI

Central Limit Theorem (CLT) - Example 2

$n = 1 \qquad Y_1 = \frac{X_1}{1}$

Create many samples of Y_1 so you can get a pretty histogram



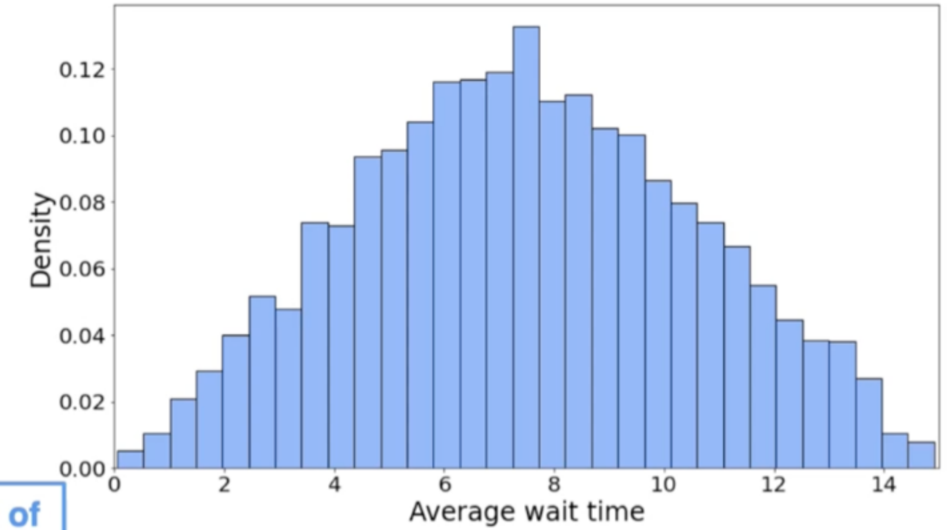
What happens to the distribution of these averages as n increases?

DeepLearning.AI

Central Limit Theorem (CLT) - Example 2

$n = 2 \qquad Y_2 = \frac{X_1 + X_2}{2}$

Create many samples of Y_2 so you can get a pretty histogram



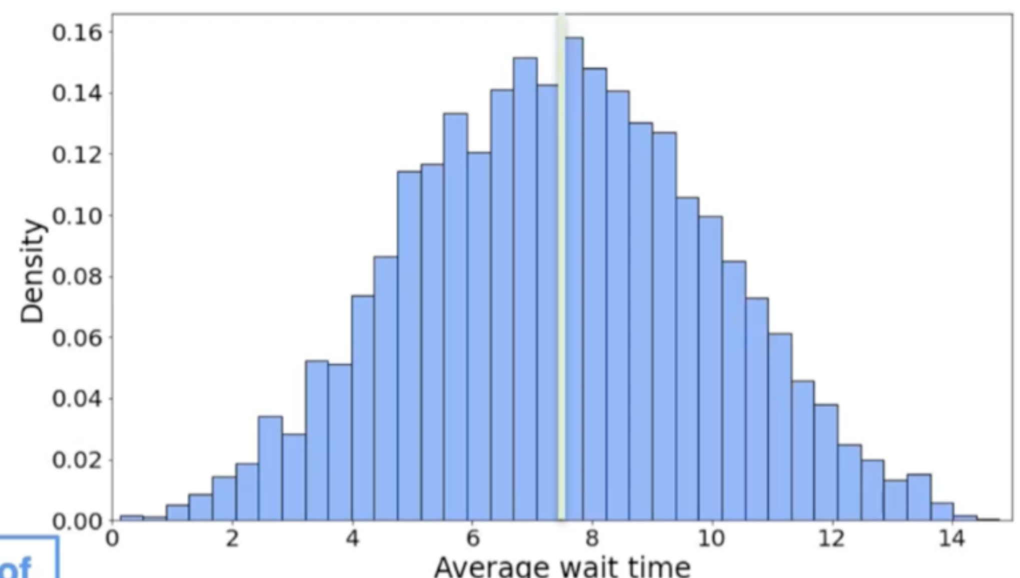
What happens to the distribution of these averages as n increases

DeepLearning.AI

Central Limit Theorem (CLT) - Example 2

$n = 3 \quad Y_3 = \frac{X_1 + X_2 + X_3}{3}$

Create many samples of Y_3 so you can get a pretty histogram



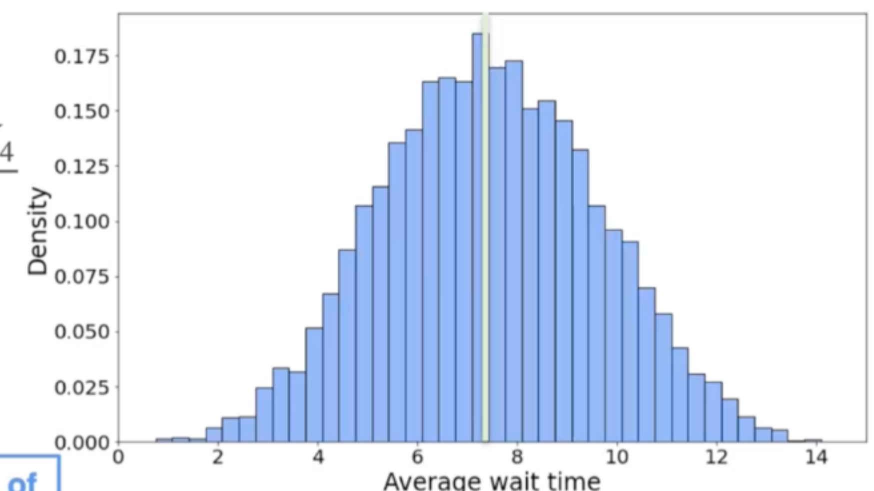
What happens to the distribution of these averages as n increases

DeepLearning.AI

Central Limit Theorem (CLT) - Example 2

$n = 4 \quad Y_4 = \frac{X_1 + X_2 + X_3 + X_4}{4}$

Create many samples of Y_4 so you can get a pretty histogram



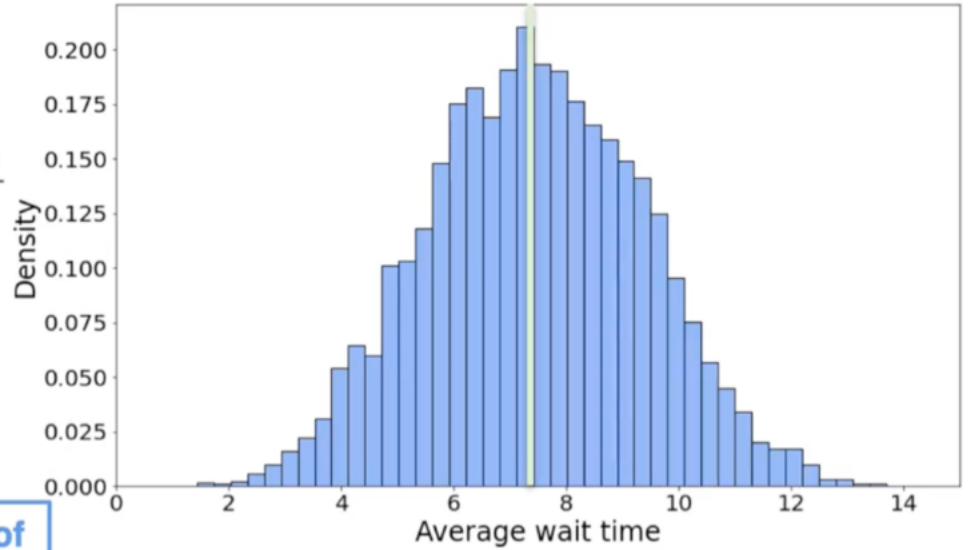
What happens to the distribution of these averages as n increases

DeepLearning.AI

Central Limit Theorem (CLT) - Example 2

$n = 5 \quad Y_5 = \dfrac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$

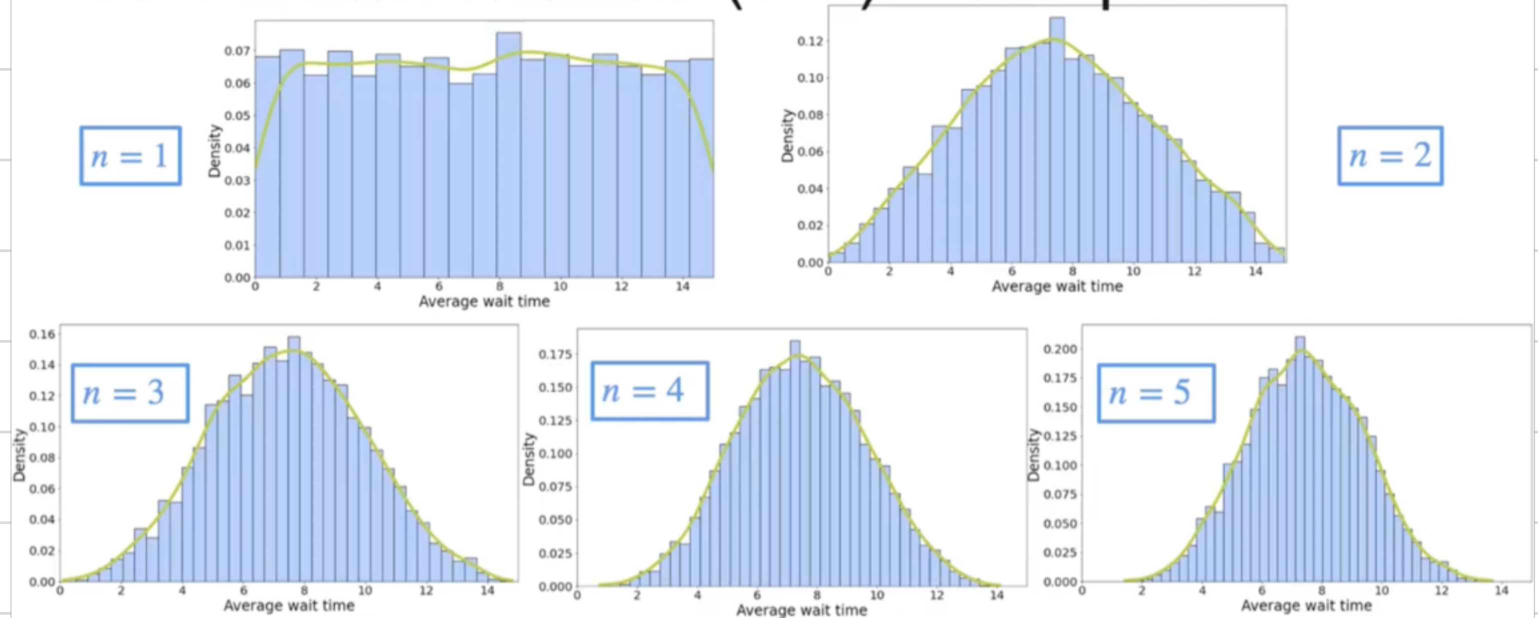
Create many samples of Y_5 so you can get a pretty histogram



What happens to the distribution of these averages as n increases

DeepLearning.AI

Central Limit Theorem (CLT) - Example 2



DeepLearning.AI

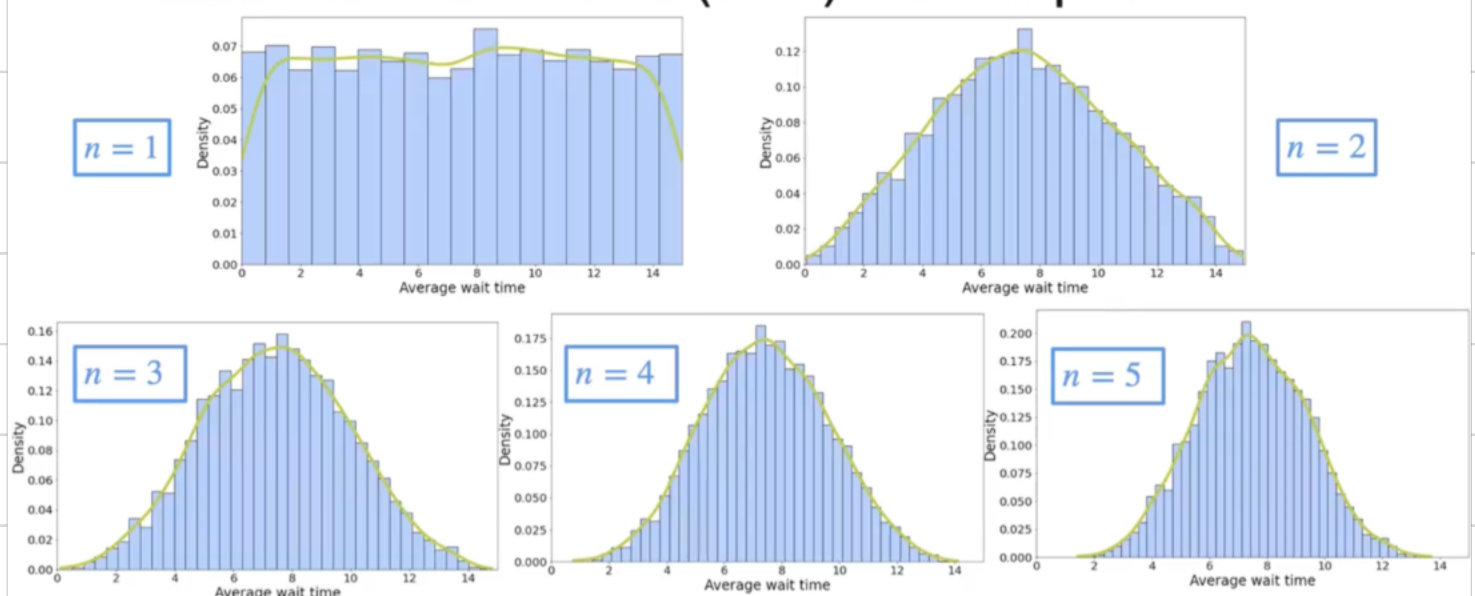
Central Limit Theorem (CLT) - Example 2

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n} n\mathbb{E}[X] = \mathbb{E}[X] = 7.5$$

$$Var[Y_n] = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$
$$= \frac{1}{n^2} n Var(X) = \frac{Var(X)}{n}$$

DeepLearning.AI

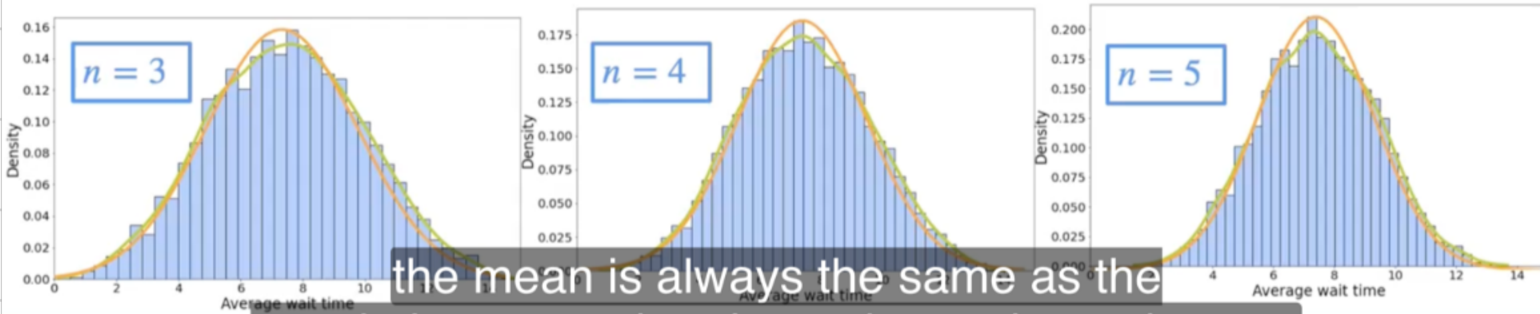
Central Limit Theorem (CLT) - Example 2



DeepLearning.AI

Central Limit Theorem (CLT) - Example 2

$$\frac{Y_n - 7.5}{\sqrt{18.75/n}}$$



the mean is always the same as the population mean, but the variance depends on n.

Central Limit Theorem (CLT) - Example 2

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\text{mean of } Y_n \rightarrow \mu_{Y_n} = \mu \leftarrow \text{population mean}$$

$$\text{variance of } Y_n \rightarrow \sigma^2_{Y_n} = \frac{\sigma^2}{n} \leftarrow \text{population variance}$$

so there is less spread and less variance.

Central Limit Theorem (CLT) - Formal Definition

As $n \to \infty$ $$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]}{\sigma_X}\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$$

Central Limit Theorem (CLT) - Formal Definition

As $n \to \infty$ $$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]}{\sigma_X}\sqrt{n} \sim \mathcal{N}\left(0,1^2\right)$$

As $n \to \infty$ $$\frac{1}{\cancel{n}\ \sqrt{n}}\left(\frac{\sum_{i=1}^{n} X_i - \frac{1}{\cancel{n}}\cancel{n}\mathbb{E}[X]}{\sigma_X}\right)\cancel{\sqrt{n}} \sim \mathcal{N}\left(0,1^2\right)$$

factor 1 over n and rearranging terms properly.