deeplearning.ai

Evaluation Metrics

Lesson 1

How good is a model?

$$\text{Accuracy} = \frac{\text{Examples correctly classified}}{\text{Total number of examples}}$$

Ground Truth	Model 1
Normal	-
Normal	-
Normal	-
Normal	-
Normal	-
Disease	-
Normal	-
Disease	-
Normal	-
Normal	-

$\frac{8}{10}$

1.00

$\frac{8}{10} = 0.8$

Ground Truth	Model 2	Model 1
Normal	-	-
Normal	+	-
Normal	-	-
Normal	-	-
Normal	-	-
Disease	+	-
Normal	-	-
Disease	+	-
Normal	+	-
Normal	-	-

$$\text{Accuracy} = P(\text{correct})$$

$$\text{Accuracy} = P(\text{correct} \cap \text{disease}) + P(\text{correct} \cap \text{normal})$$

Using $P(A \cap B) = P(A \mid B)\,P(B)$

$$\text{Accuracy} = P(\text{correct} \mid \text{disease})P(\text{disease}) + P(\text{correct} \mid \text{normal})P(\text{normal})$$

$$\text{Accuracy} = P(+ \mid \text{disease})P(\text{disease}) + P(- \mid \text{normal})P(\text{normal})$$

$$\text{Accuracy} = P(\text{correct})$$

$$\text{Accuracy} = P(\text{correct} \cap \text{disease}) + P(\text{correct} \cap \text{normal})$$

Using $P(A \cap B) = P(A \mid B)\,P(B)$

$$\text{Accuracy} = P(\text{correct} \mid \text{disease})P(\text{disease}) + P(\text{correct} \mid \text{normal})P(\text{normal})$$

$$\text{Accuracy} = P(+ \mid \text{disease})P(\text{disease}) + P(- \mid \text{normal})P(\text{normal})$$

Sensitivity (true positive rate) Specificity (true negative rate)

P(+ | disease)

If a patient has the disease, what is the probability that the model predicts positive?

Sensitivity

P(− | normal)

If a patient is normal, what is the probability that the model predicts negative?

Specificity

1.00

Share

Accuracy = P(correct)

Accuracy = Sensitivity × P(disease) + Specificity × P(normal)

1 - P(disease)

Prevalence

Accuracy = Sensitivity × prevalence + Specificity × (1 − prevalence)

1:18 / 4:08

1.00

Share

$$\text{Accuracy} = \text{Sensitivity} \times \text{prevalence} + \text{Specificity} \times (1 - \text{prevalence})$$

1:49 / 4:08

Ground Truth	Model
Normal	-
Normal	-
Disease	+
Normal	-
Normal	-
Disease	-
Normal	-
Disease	+
Normal	+
Normal	-

Sensitivity

P(+ | disease)

$$\frac{\#(+ \text{ and disease})}{\#(\text{disease})} = \frac{2}{3} = 0.67$$

Specificity

P(− | normal)

$$\frac{\#(- \text{ and normal})}{\#(\text{normal})} = \frac{6}{7} = 0.86$$

1.00

Ground Truth	Model
Normal	-
Normal	-
Disease	+
Normal	-
Normal	-
Disease	-
Normal	-
Disease	+
Normal	+
Normal	-

Sensitivity = 0.67

Specificity = 0.86

Prevalence

P(*disease*)

$$\frac{\#(\text{disease})}{\#(\text{total})} = \frac{3}{10} = 0.3$$

Accuracy

$$\text{Sensitivity} \times \text{prevalence} + \text{Specificity} \times (1 - \text{prevalence})$$

$$= 0.67 \times 0.3 + 0.86 \times 0.7 = 0.8$$

1.00

Sensitivity

$$P(\underline{+} \mid \underline{\text{disease}})$$

If a patient has the disease, what is the probability that the model predicts positive?

$\downarrow$

If a model prediction is positive, what is the probability that a patient has the disease?

$$P(\underline{\text{disease}} \mid \underline{+})$$

PPV

1.00

Specificity

$$P(\underline{-} \mid \underline{\text{normal}})$$

If a patient is normal, what is the probability that the model predicts negative?

↓

If a model prediction is negative, what is the probability that a patient is normal?

$$P(\underline{\text{normal}} \mid \underline{-})$$

NPV

Ground Truth	Model
Normal	-
Disease	+
Normal	+
Normal	-
Normal	-
Disease	-
Normal	-
Disease	+
Normal	+
Normal	-

PPV

$$P(\text{disease} \mid +)$$

$$\frac{\#(+\text{ and disease})}{\#(+)} = \frac{2}{4} = 0.5$$

NPV

$$P(\text{normal} \mid -)$$

$$\frac{\#(-\text{ and normal})}{\#(-)} = \frac{5}{6} = 0.83$$

$$P(\text{disease} \mid +)$$

PPV

$$P(\text{normal} \mid -)$$

NPV

$$P(+ \mid \text{disease})$$

Sensitivity

$$P(- \mid \text{normal})$$

Specificity

1.00

Ground Truth	Model
Normal	-
Disease	+
Normal	+
Normal	-
Normal	-
Disease	-
Normal	-
Disease	+
Normal	+
Normal	-

Model Output

GT	+	-
Disease	2	1
Normal	2	5

Model Output

		+	-
GT	Disease	True Positive (TP)	False Negative (FN)
	Normal	False Positive (FP)	True Negative (TN)

$$\frac{\#(+\text{ and disease})}{\#(\text{disease})} = \text{Sensitivit}$$

$$\frac{\#(-\text{ and normal})}{\#(\text{normal})} = \text{Specificity}$$

$$\frac{\#(+\text{ and disease})}{\#(+)} = \text{PPV}$$

$$\frac{\#(-\text{ and normal})}{\#(-)} = NPV$$

1.00

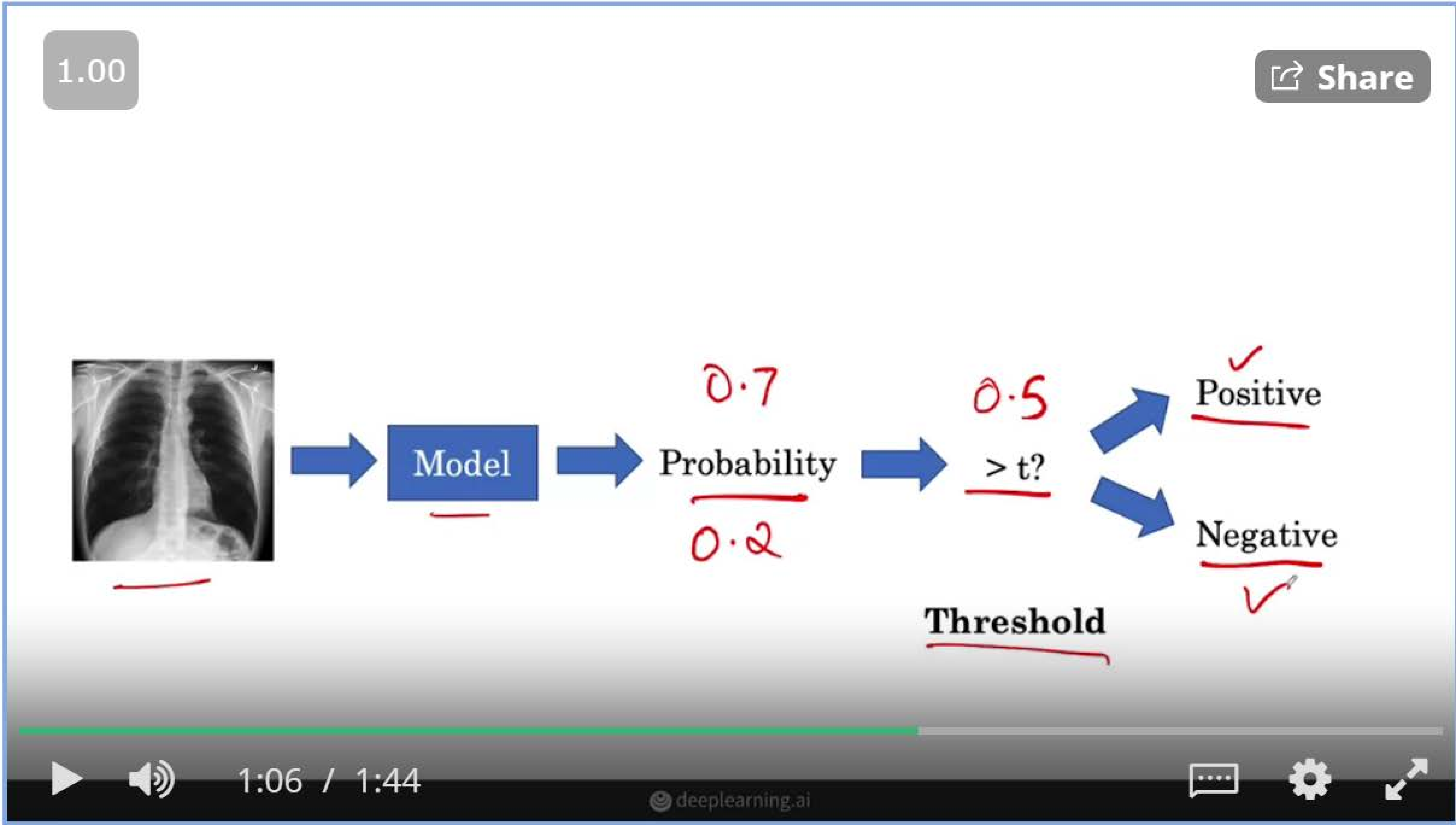
Model Output

GT		+	-
	Disease	True Positive (TP)	False Negative (FN)
	Normal	False Positive (FP)	True Negative (TN)

$$\frac{TP}{TP+FN} = \text{Sensitivity}$$

$$\frac{TN}{FP+TN} = \text{Specificity}$$

$$\frac{TP}{TP+FP} = \text{PPV} \qquad \frac{TN}{TN+FN} = \text{NPV}$$


1.00
Share
Model
0·7
Probability
0·2
0·5
> t?
Positive
Negative
Threshold
1:06 / 1:44
deeplearning.ai

1.00

Share

$$P(+ \mid \text{disease}) \qquad P(- \mid \text{normal})$$

Sensitivity **Specificity**

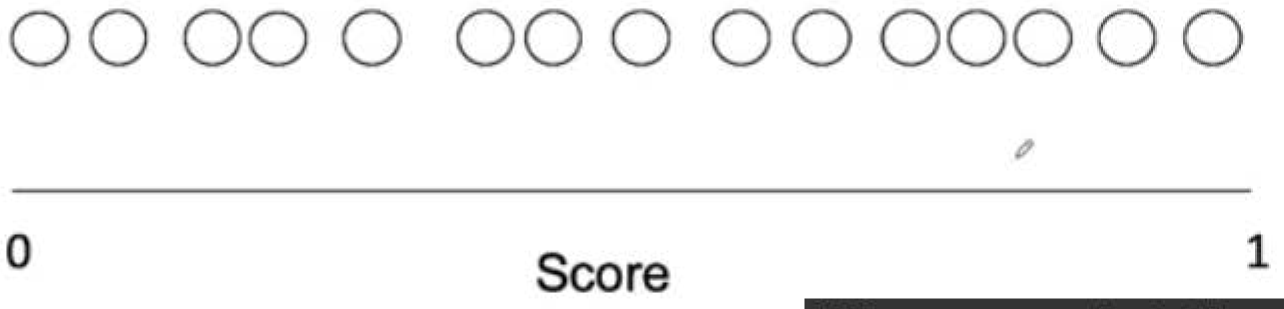
$t = 0$

$t = 1$

1:38 / 1:44

1.00

X-Ray	Output Probability (Score)
1	0.30
2	0.42
3	0.78
…	…
15	0.98



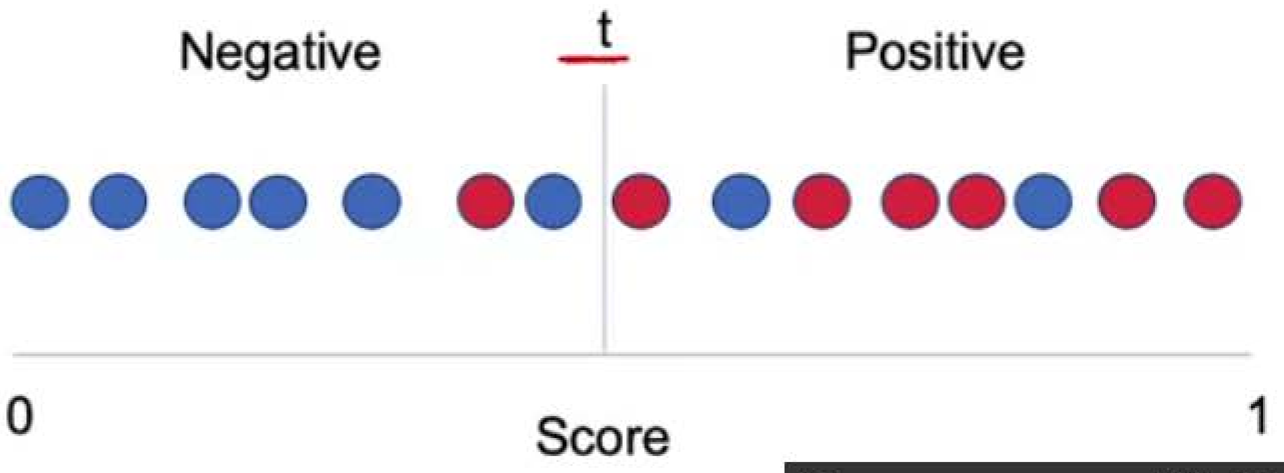
We can plot these
15 outputs scores

1.00

$$P(+ \mid \text{disease}) \qquad P(- \mid \text{normal})$$

Sensitivity **Specificity**

$\frac{6}{7} = 0.85$ $\frac{6}{8}$



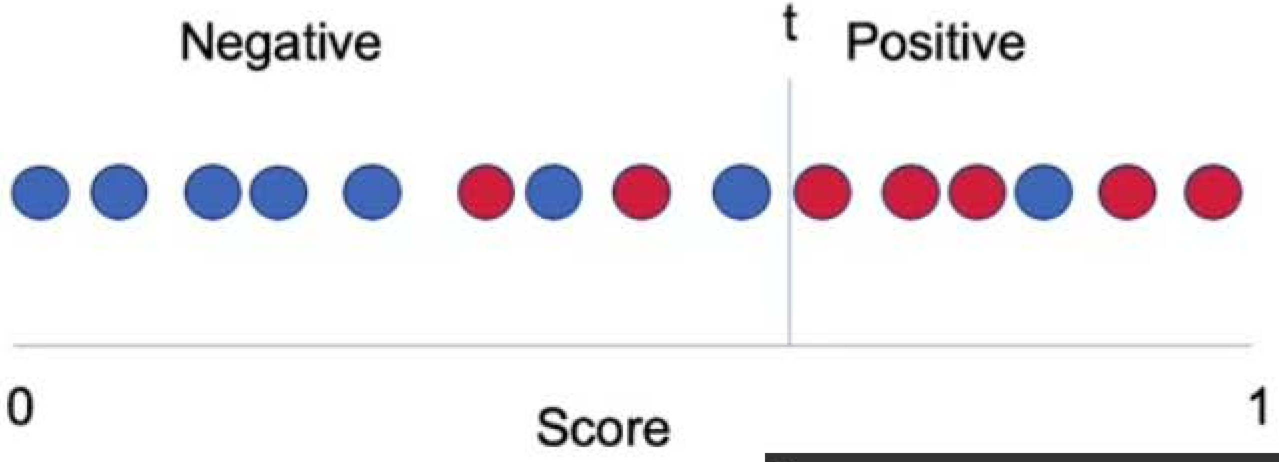
So our specificity is
six over eight or 0.75.

1.00

$$P(+ \mid \text{disease}) \qquad P(- \mid \text{normal})$$

Sensitivity **Specificity**

$$\frac{5}{7} = 0.71 \qquad \frac{7}{8} = 0.88$$



because we are now
correctly classifying

1.00

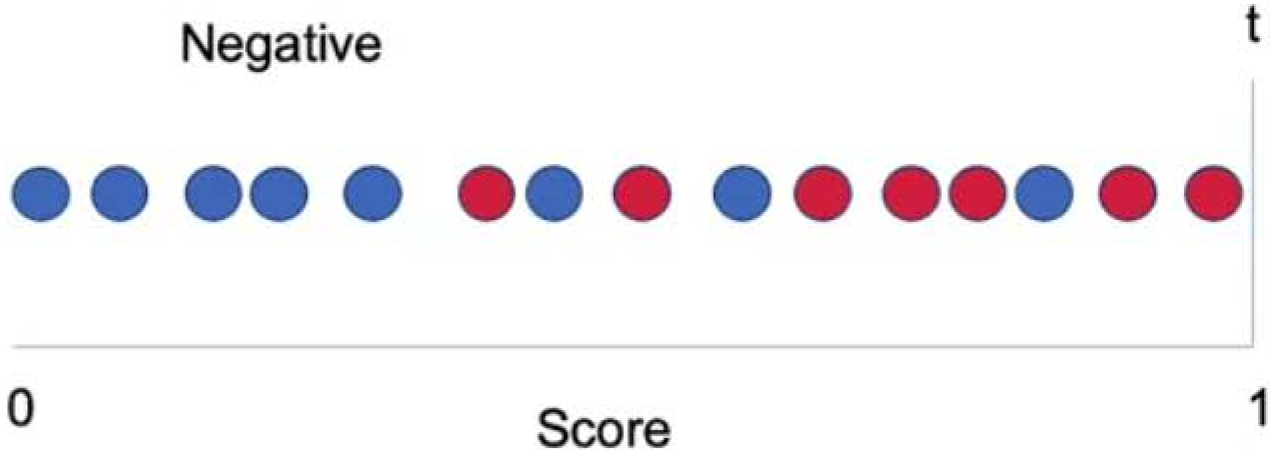
$$P(+ \mid \text{disease}) \qquad P(- \mid \text{normal})$$

Sensitivity

$$\frac{0}{7} = 0$$

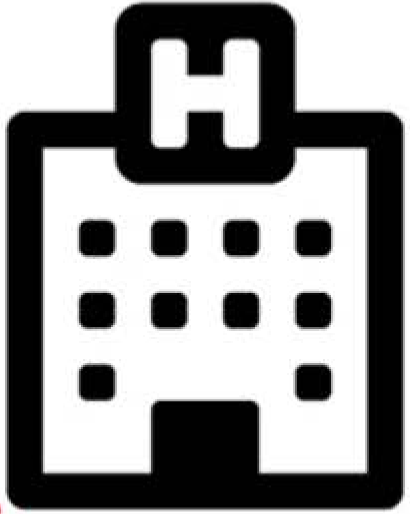
Specificity

$$\frac{8}{8} = 1$$



since all the examples are
classified as negative.

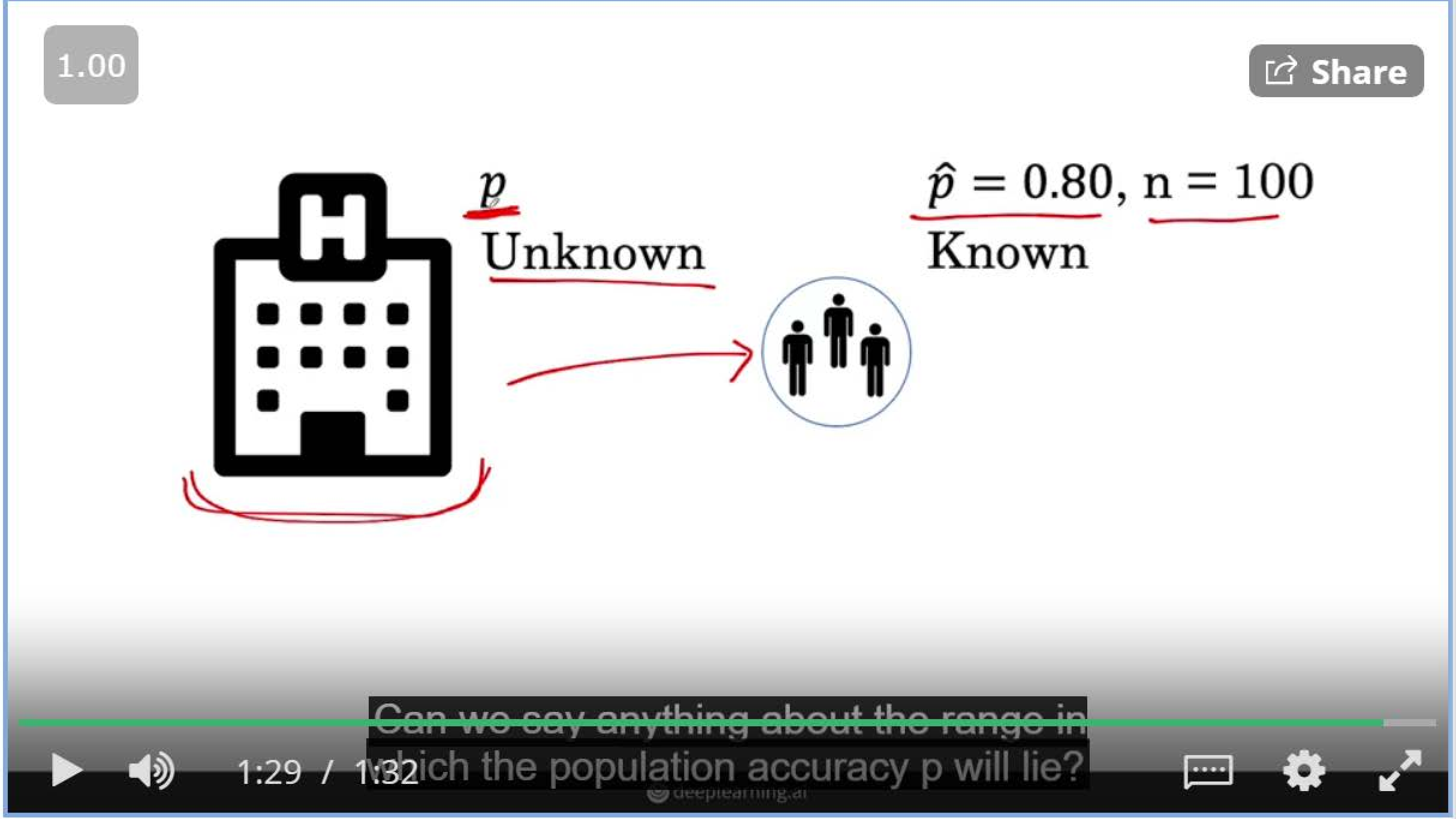
1.00

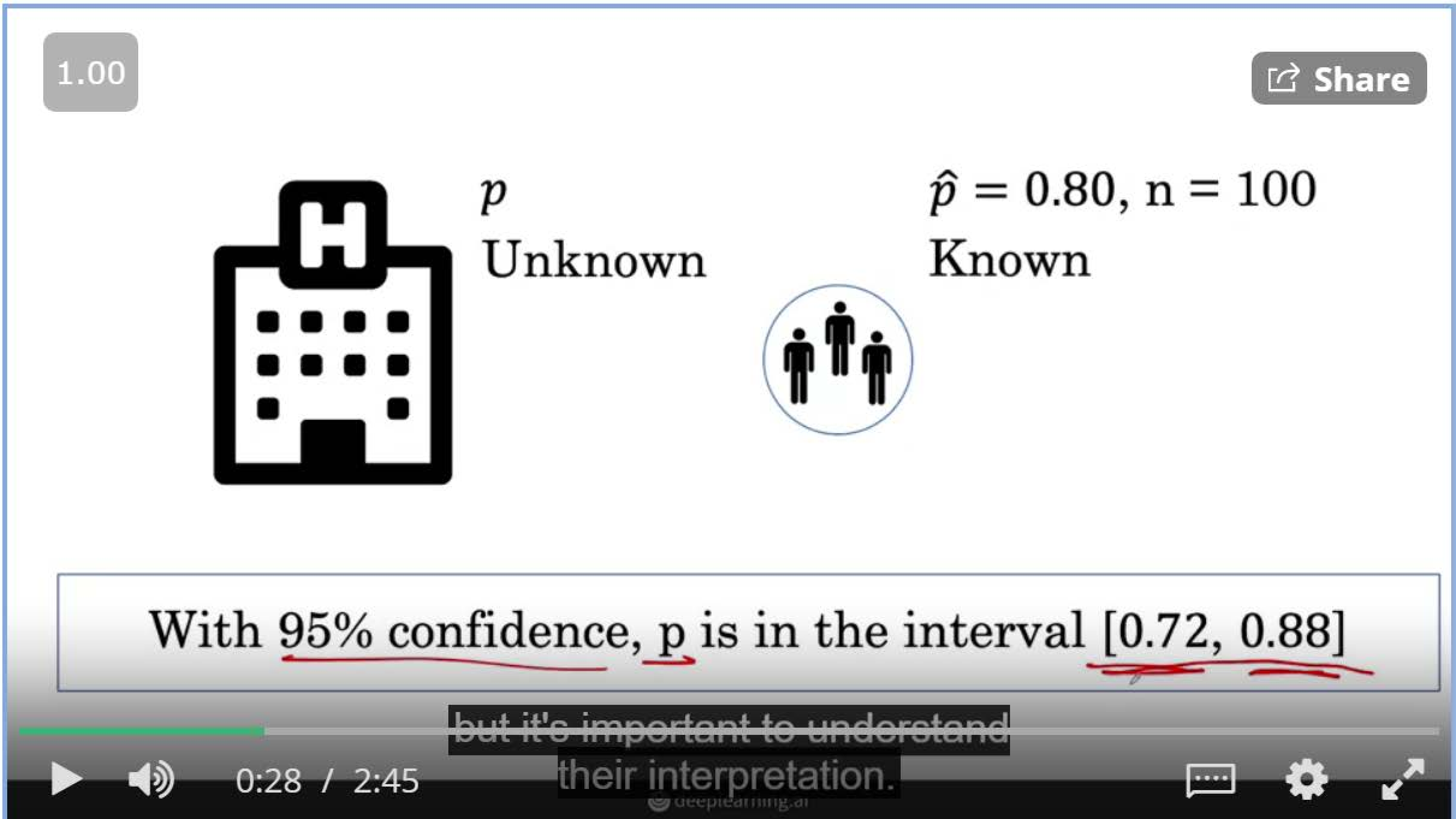


$p = 0.78$

50000 patients

This is called the population accuracy,
here with small p.


1.00
Share
p
Unknown
$\hat{p} = 0.80$, n = 100
Known
Can we say anything about the range in
which the population accuracy p will lie?
1:29 / 1:32
deeplearning.ai


1.00
Share
p
Unknown
$\hat{p} = 0.80$, n = 100
Known
With 95% confidence, p is in the interval [0.72, 0.88]
but it's important to understand
their interpretation.
0:28 / 2:45

0.80 (95% CI 0.72, 0.88)

Interpretation of 95% confidence interval

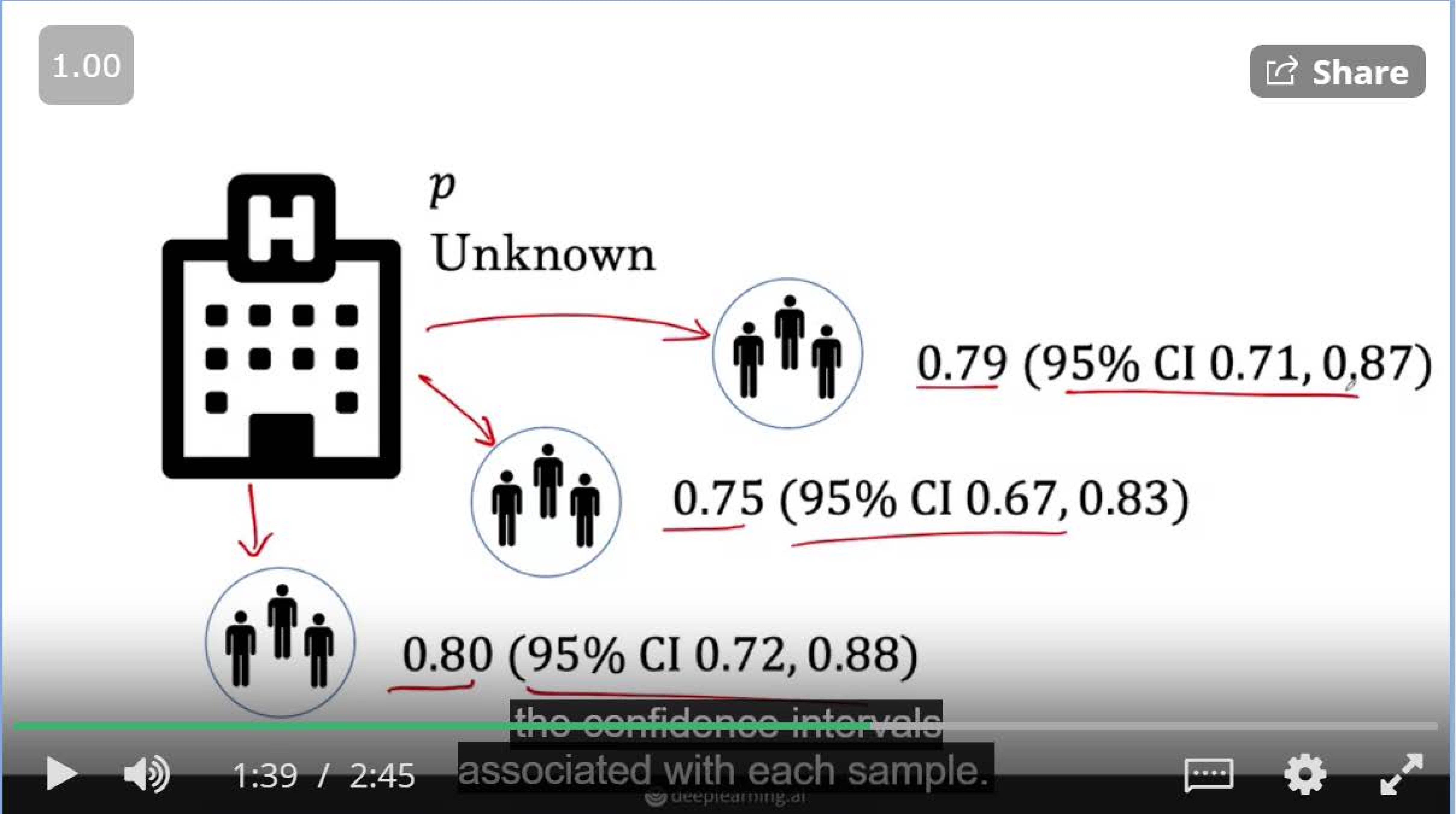
With 95% confidence, p is in the interval [0.72, 0.88]

Mis-interpretation

There is a 95% probability that p lies within the interval [0.72, 0.88]

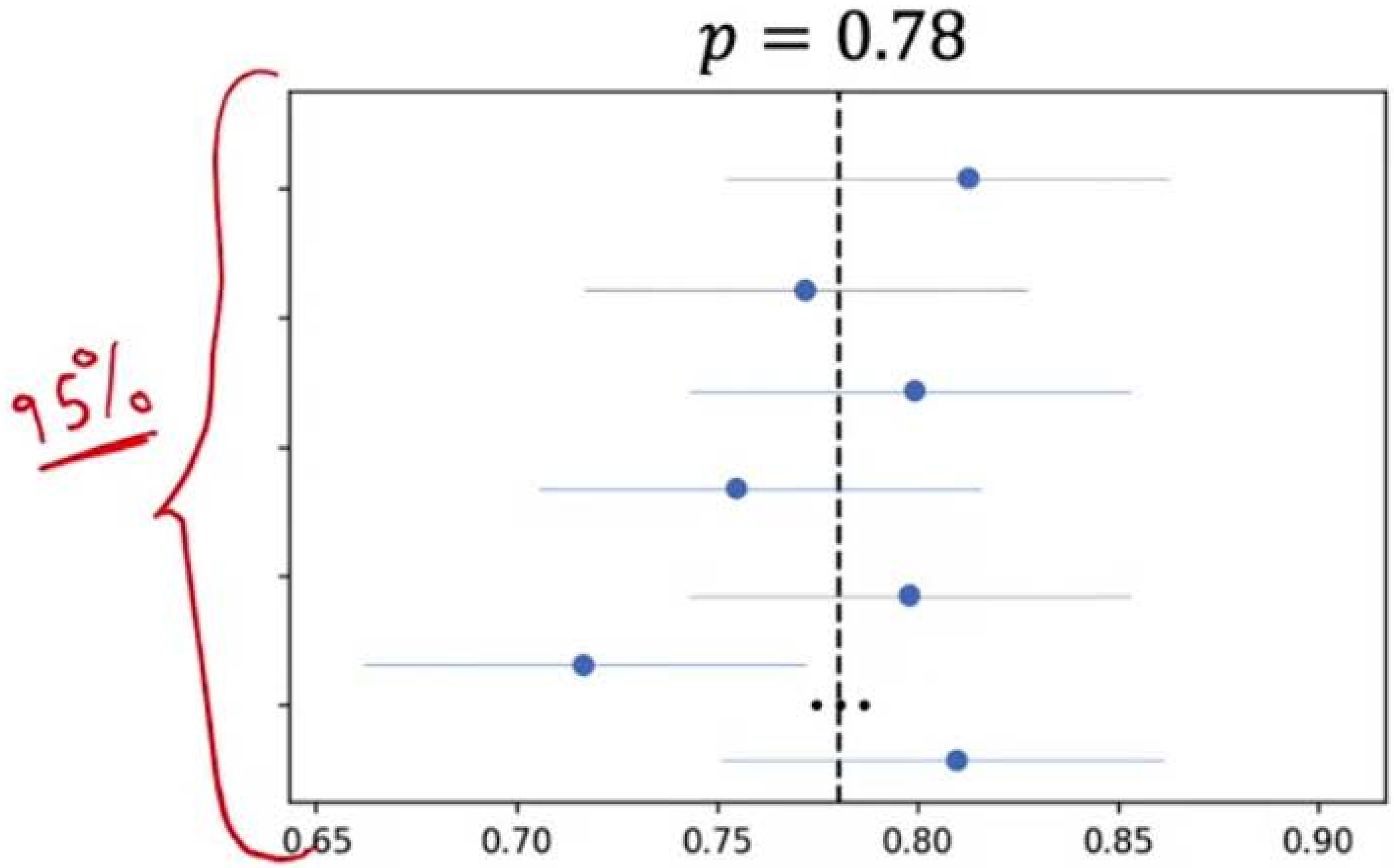
Mis-interpretation

95% of the sample accuracies lie within the interval [0.72, 0.88]


1.00
Share
p
Unknown
0.79 (95% CI 0.71, 0,87)
0.75 (95% CI 0.67, 0.83)
0.80 (95% CI 0.72, 0.88)
the confidence intervals
associated with each sample.
1:39 / 2:45

1.00

Share



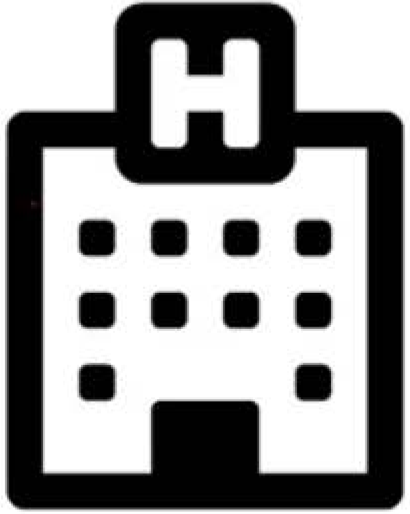
Interpretation of 95% confidence

In repeated sampling, this method produces intervals that include the population accuracy in about 95% of samples

Play

2:43 / 2:45

1.00



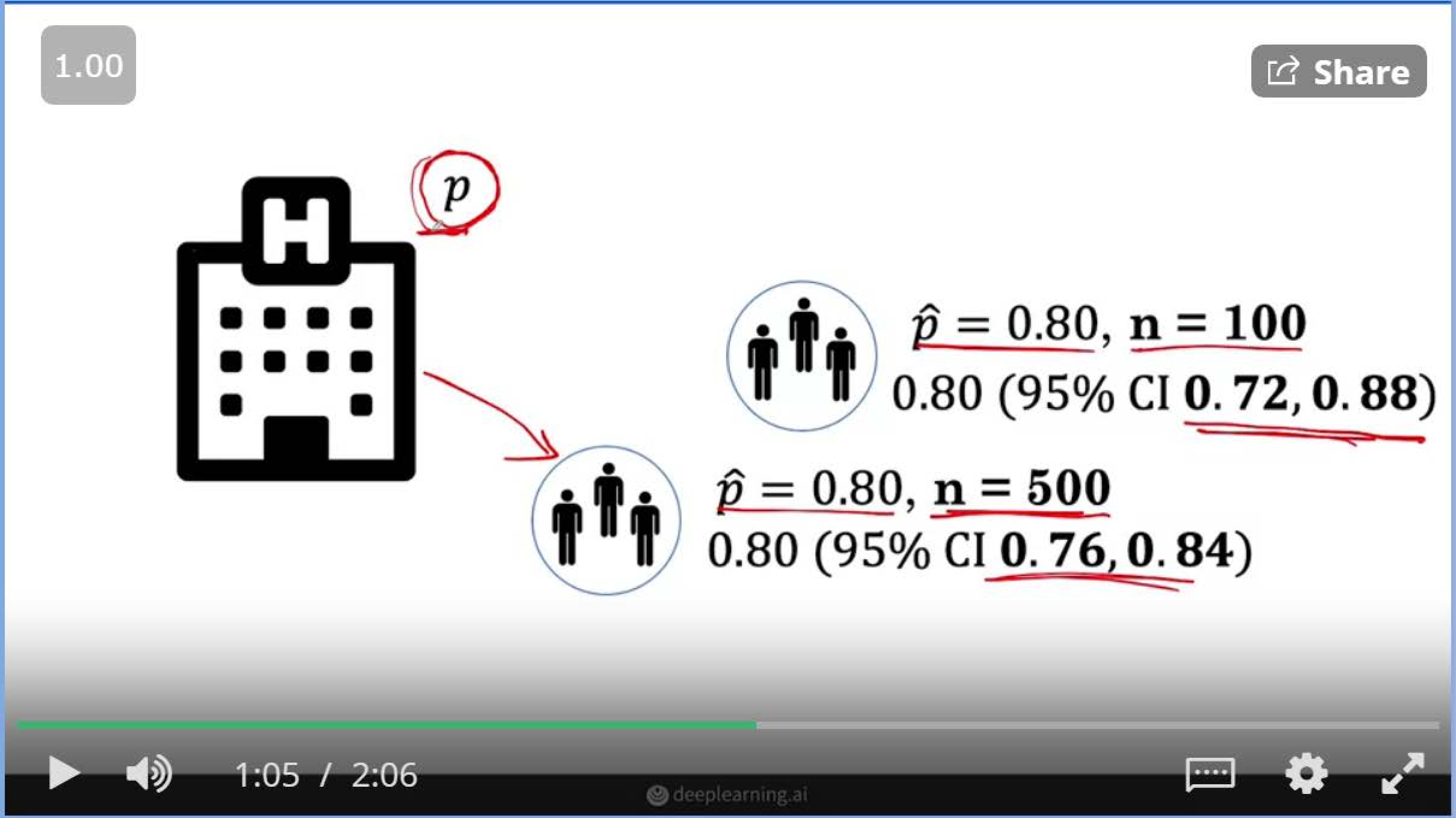
p

Unknown

$\hat{p} = 0.80$, n = 100

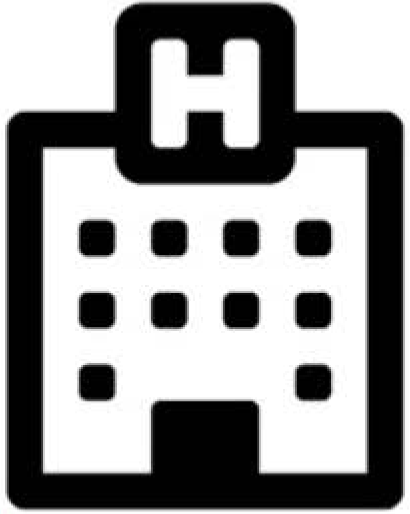
0.80 (95% CI 0.72, 0.88)

Known



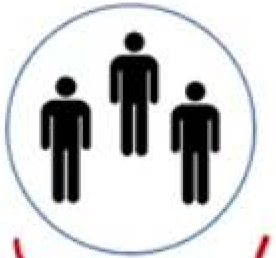
1.00
Share
p
$\hat{p} = 0.80$, $\mathbf{n = 100}$
0.80 (95% CI $\mathbf{0.72, 0.88}$)
$\hat{p} = 0.80$, $\mathbf{n = 500}$
0.80 (95% CI $\mathbf{0.76, 0.84}$)
1:05 / 2:06
deeplearning.ai

1.00



p

Unknown



0.80 (95% CI $0.72, 0.88$)

Known