Day-68, Feb 06, 2024 (Magh 24, 2081 B.S.)

1. t-tests, t-statistics, t-distribution, Normal Distribution
2. Hypothesis Testing → Power-test
3. Hypothesis Testing → Intrepreting Results
4. Test for Proportions (Reading)

DeepLearning.AI

Hypothesis Testing

Power of a test

H_0, H_1.

Power of a test tells how reliable you reject the Null Hypothesis

1. Type I and Type II Errors Recap:
- Type I Error (False Positive): Rejecting the null hypothesis ($H0$) when it's actually true.
In this example, concluding that the mean height of 18-year-olds has increased when it's still 66.7 inches.
- Type II Error (False Negative): Failing to reject the null hypothesis when it's actually false.
In this case, concluding that the mean height is still 66.7 inches when it's actually greater than 66.7.

2. Significance Level (α):
You set a significance level of 0.05, meaning a 5% chance of making a Type I Error.
The critical value for rejecting the null hypothesis was determined to be 68.26.

3. Type II Error (β):
Type II Error occurs when you fail to reject H_0 when it's false.
If the true mean is 70 inches, you want to know the probability of getting a sample mean smaller than 68.26.

4. Gaussian Distribution Insight:
If the true population mean is 66.7, the sample mean follows a Gaussian distribution with:
Mean = 66.7
Standard deviation = 3/ ROOT10

If the true population mean is 70, the distribution shifts to:
Mean = 70
Same standard deviation (3/ROOT 10).

5. Probability of Type II Error at μ =70:
You are looking for the probability that the sample mean is smaller than 68.26 when the population mean is 70.
This shaded blue region is calculated as β= 0.33.

6.Power of the Test (1−β):
The power of the test is the probability of correctly rejecting the null hypothesis when it is false.

Power = 1−β.
In this case, Power = 1−0.33=0.67.
This tells you that if the true mean is 70 inches, there's a 67% chance you'll correctly reject the null hypothesis.

7. Power Curve Explanation:
The power curve plots the probability of rejecting H_0 for different values of the true population mean (μ).
At μ=66.7, the power is exactly α (in this case, 0.05).
As μ increases beyond 66.7, the curve rises, approaching 1 because it becomes easier to reject H_0.

8. Impact of Changing α:
Power curves for α =0.01,0.05,0.1 show that increasing α increases the power.
This makes sense: when you allow more room for Type I errors, you're less likely to miss true effects (lower Type II errors).

9. Trade-off Between Type I and Type II Errors:
If you decrease α (become stricter about Type I errors), you increase β (more Type II errors).
There's always a trade-off between these errors unless you can increase the sample size.

Summary:
- Type I Error happens when you mistakenly think there's a change.
- Type II Error happens when you mistakenly think there's no change.

The Power of the Test measures how likely you are to correctly detect a true change.
There's a trade-off between Type I and Type II errors, but a larger sample size can reduce both.

Power of a Test

The power of a test is the probability of correctly rejecting the null hypothesis (H_0) when it is false. In other words, it measures how likely the test is to detect a real effect or difference.

Formula:

Power=1−β Where: β is the probability of a Type II error (failing to reject H_0 when it is false).

Why It Matters:

A higher power means your test is more reliable in detecting true changes or effects, reducing the chance of false negatives.

Factors Affecting Power:

Sample Size: Larger samples increase power.

Significance Level (α): Higher α increases power but raises Type I error risk.

Effect Size: Larger differences between true and null values increase power.

Variability: Lower data variability increases power.

Concise Summary:

Type I Error: Rejecting the null hypothesis (H_0) when it's true (false positive).

Type II Error: Failing to reject H_0 when it's false (false negative).

In this example:H_0: The mean height is still 66.7 inches. Critical value: 68.26 (reject if the sample mean > 68.26).

Type II Error at μ=70: Probability (β) is 0.33.

Key Concepts:

Power of Test (1−β): Probability of correctly rejecting H_0. Here, 0.67 at μ=70.

Higher α (significance level) increases the power but also Type I errors.

There's a trade-off between Type I and Type II errors; larger sample sizes reduce both.

Type I and Type II Errors

1

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Decision	Reality	
	H_0 True ($\mu = 66.7$)	H_0 False ($\mu > 66.7$)
Reject H_0 (Decide $\mu > 66.7$)	Type I error	Correct
Don't reject H_0 (Decide $\mu = 66.7$)	Correct	Type II error

one value of the population mean, in this case 66.7.

DeepLearning.AI

Type I and Type II Errors

2

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Decision	Reality	
	H_0 True ($\mu = 66.7$)	H_0 False ($\mu > 66.7$)
Reject H_0 (Decide $\mu > 66.7$)	Type I error	Correct
Don't reject H_0 (Decide $\mu = 66.7$)	Correct	Type II error

DeepLearning.AI However, type II errors can happen for

Finding the Type II Error Probabilities

3

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$ $\quad n = 10 \quad \sigma = 3$

For $\alpha = 0.05$: $k_\alpha = 68.26$ Decision rule: Reject H_0 if $\bar{x} > 68.26$

What is the **Type II error probability** if the true value is $\mu = 70$?

$\mathbf{P}$ (Do not reject $H_0 \mid \mu = 70) \longrightarrow \mathbf{P}(\bar{X} < 68.26 \mid \mu = 70)$

mean is smaller than 68.26, when the population mean is 70.

DeepLearning.AI

Finding the Type II Error Probabilities

4

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$ $\quad n = 10 \quad \sigma = 3$

For $\alpha = 0.05$: $k_\alpha = 68.2604$ Decision rule: Reject H_0 if $\bar{X} > 68.26$

What is the **Type II error probability** if the true value is $\mu = 70$?

If $\mu = 70 \quad \bar{X} \sim \mathcal{N}\left(70, \frac{3^2}{10}\right)$

$\beta = \mathbf{P}(\bar{X} < 68.26 \mid \mu = 70) \quad = 0.0333$

For values smaller than the critical value of 68.26, which has a value of 0.33.

DeepLearning.AI

Finding the Type II Error Probabilities

5

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

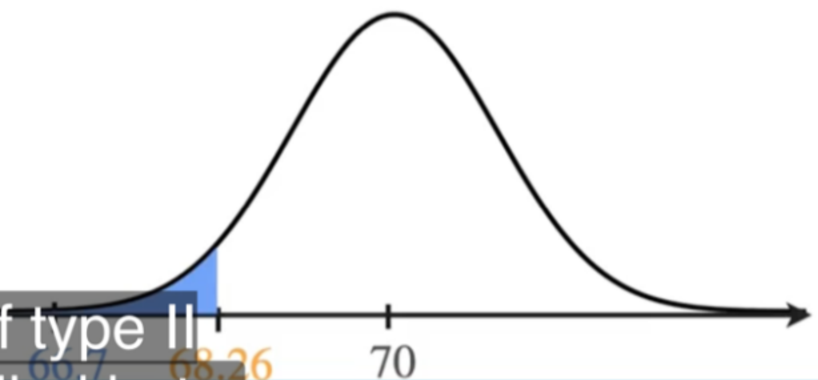
$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$ $\quad n = 10 \quad \sigma = 3$

For $\alpha = 0.05$: $k_\alpha = 68.2604$ $\quad$ Decision rule: Reject H_0 if $\overline{X} > 68.26$

What is the **Type II error probability** if the true value is $\mu = 70$?

If $\mu = 70$ $\overline{X} \sim \mathcal{N}\left(70, \frac{3^2}{10}\right)$

$\beta = \mathbf{P}(\overline{X} < 68.26 \mid \mu = 70) \quad = 0.0333$



This probability of type II errors is usually called beta.

Finding the Type II Error Probabilities

6

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$ $\quad n = 10 \quad \sigma = 3$

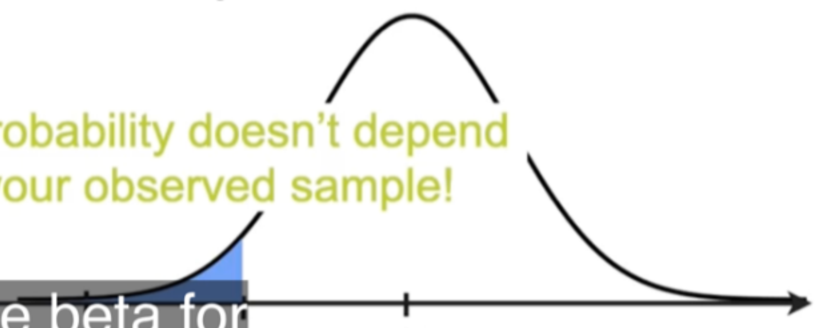
For $\alpha = 0.05$: $k_\alpha = 68.2604$ $\quad$ Decision rule: Reject H_0 if $\overline{X} > 68.26$

What is the **Type II error probability** if the true value is $\mu = 70$?

If $\mu = 70$ $\overline{X} \sim \mathcal{N}\left(70, \frac{3^2}{10}\right)$

This probability doesn't depend on your observed sample!

$\beta = \mathbf{P}(\overline{X} < 68.26 \mid \mu = 70) \quad = 0.0333$



So you can define beta for

Power of the Test

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

7

Decision	Reality: H_0 True ($\mu = 66.7$)	Reality: H_0 False ($\mu > 66.7$)
Reject H_0 (Decide $\mu > 66.7$)	Type I error	Correct
Don't reject H_0 (Decide $\mu = 66.7$	Correct	Type II error

Power of the test

$\mathbf{P}\left(\text{Reject } H_0 \,|\, \mu \in H_1\right.$

Power of the Test

8

$$\left.\begin{aligned} &\text{Type II error: } \overbrace{\mathbf{P}\left(\text{Do not reject } H_0 \,|\, \mu \in H_1\right)}^{\beta} \\ &\text{Power of the test: } \underbrace{\mathbf{P}\left(\text{Reject } H_0 \,|\, \mu \in H_1\right)}_{1-\beta} \end{aligned}\right\}$$

Complementary probabilities

To sum up for each value of mu in H1,
the power of the test

DeepLearning.AI

Power of the Test

9

$$\left.\begin{aligned} &\text{Type II error: } \overbrace{\mathbf{P}\left(\text{Do not reject } H_0 \,|\, \mu \in H_1\right)}^{\beta} \\ &\text{Power of the test: } \underbrace{\mathbf{P}\left(\text{Reject } H_0 \,|\, \mu \in H_1\right)}_{1-\beta} \end{aligned}\right\}$$

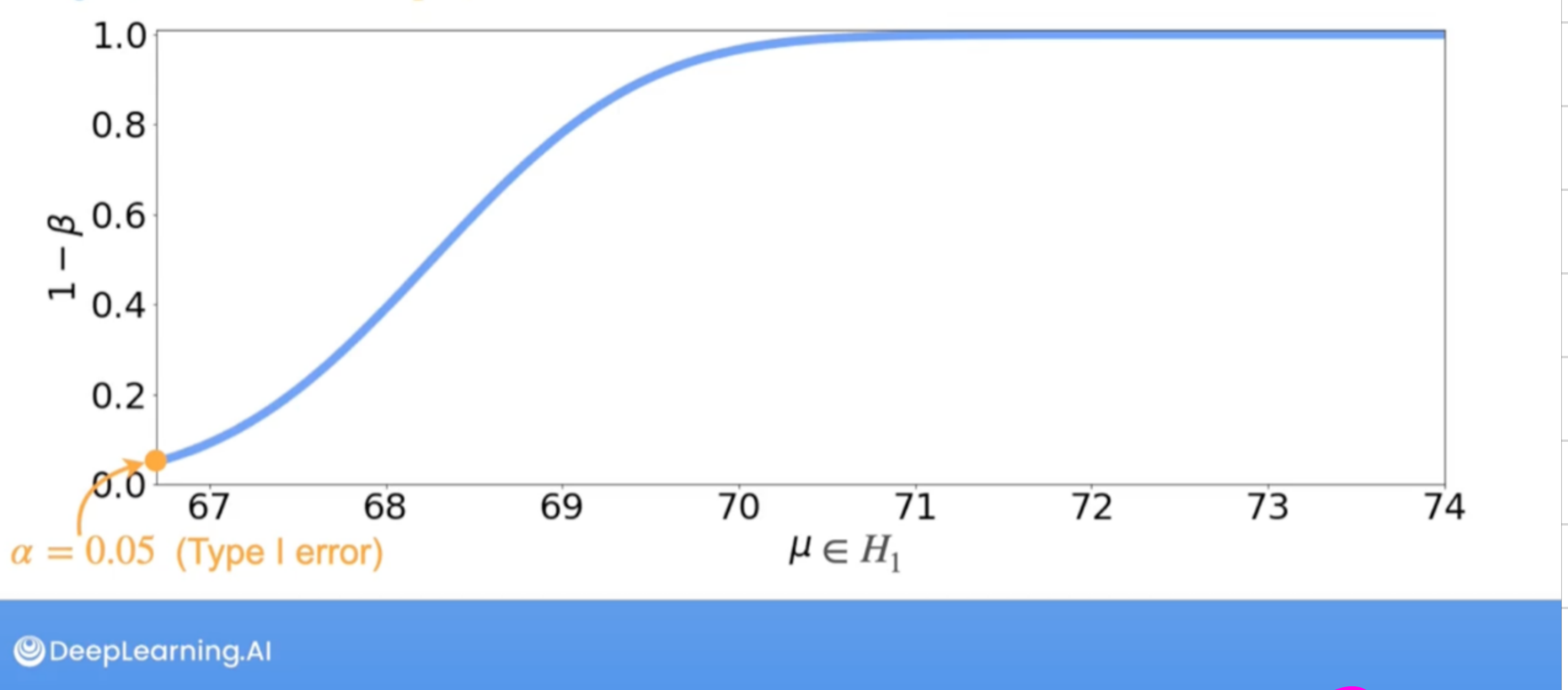
Complementary probabilities

$$\begin{aligned} \text{Power of the test } &= 1 - \text{Type II error probability} \\ &= 1 - \mathbf{P}\left(\text{Do not reject } H_0 \,|\, \mu \in H_1\right) \end{aligned}$$

DeepLearning.AI

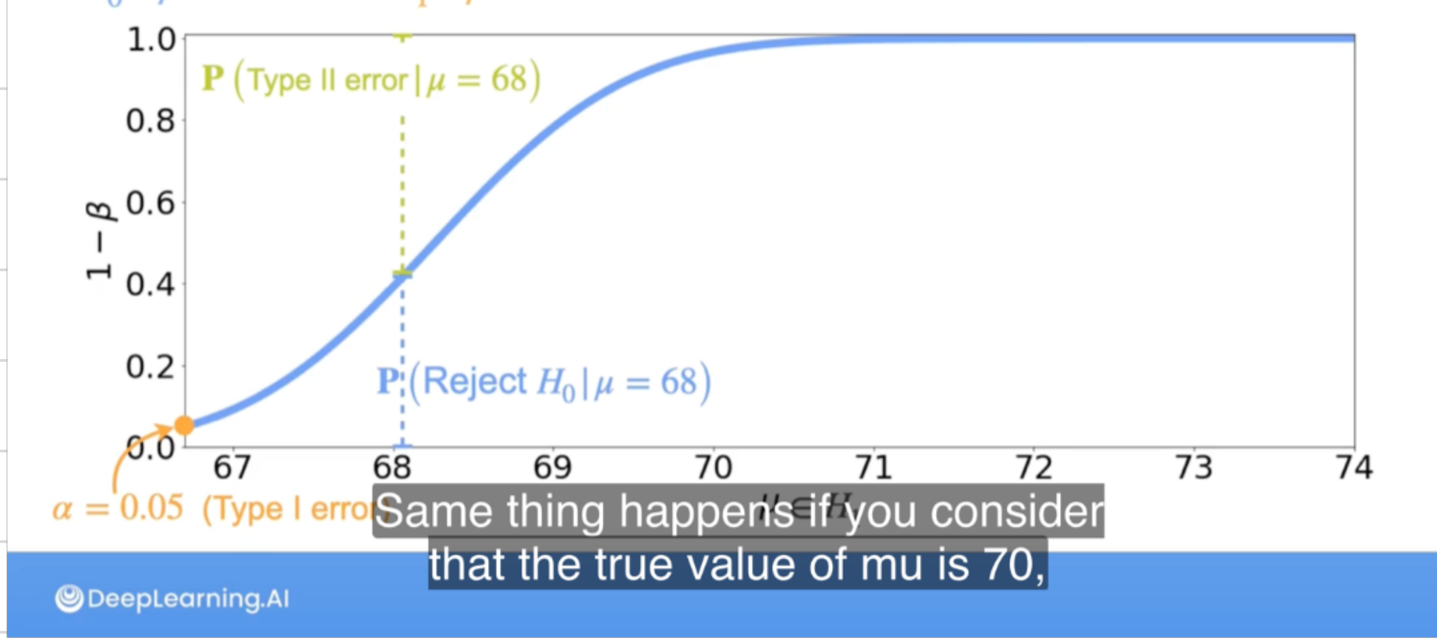
Power of the Test

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$



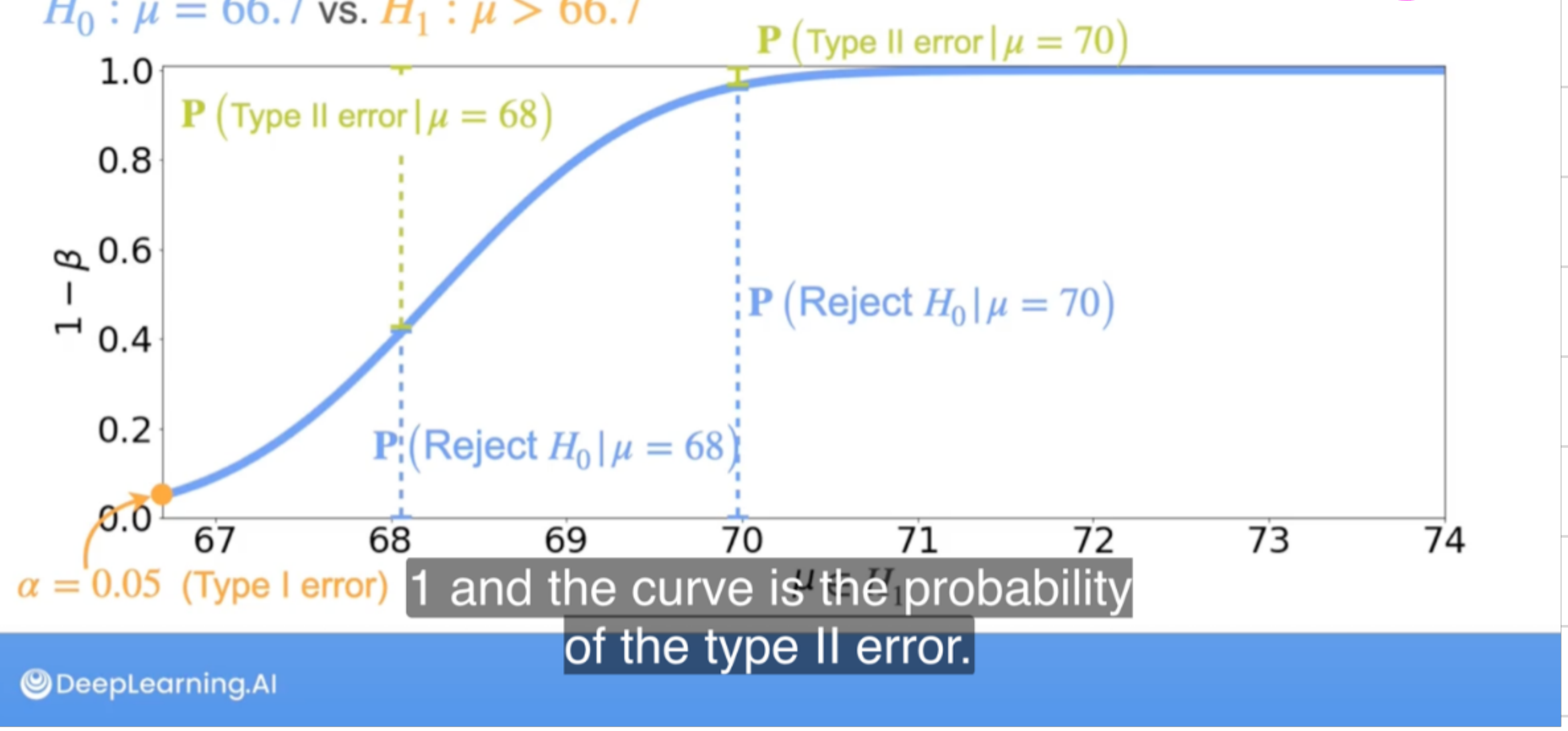
Power of the Test

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

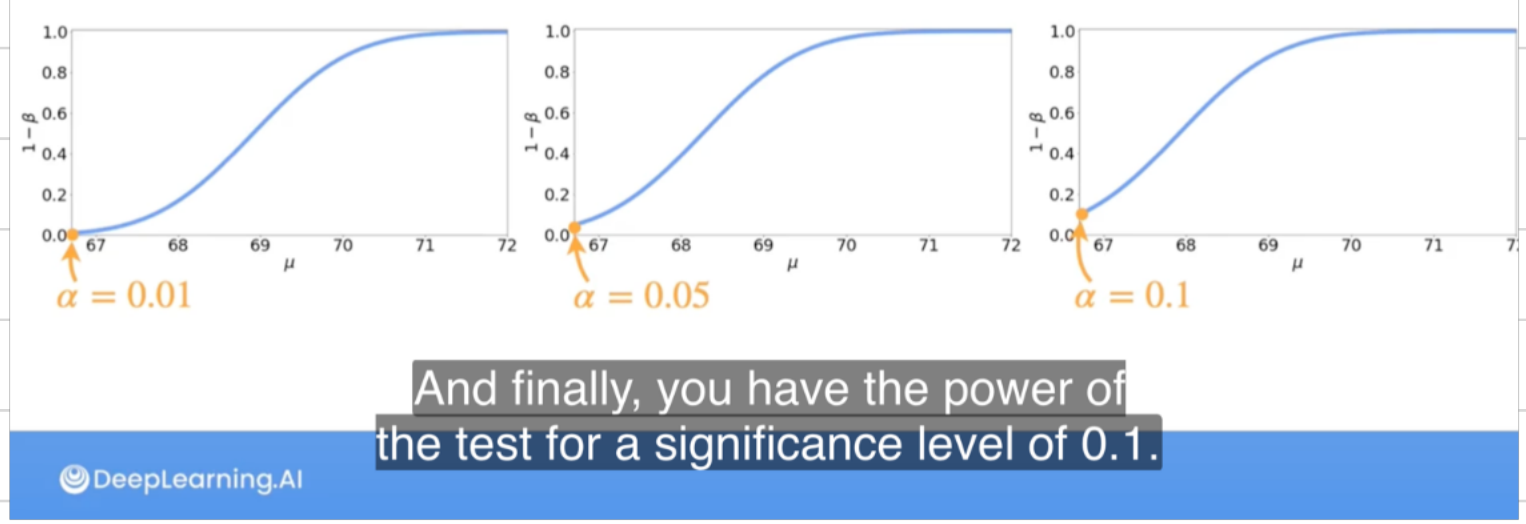


Power of the Test

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$



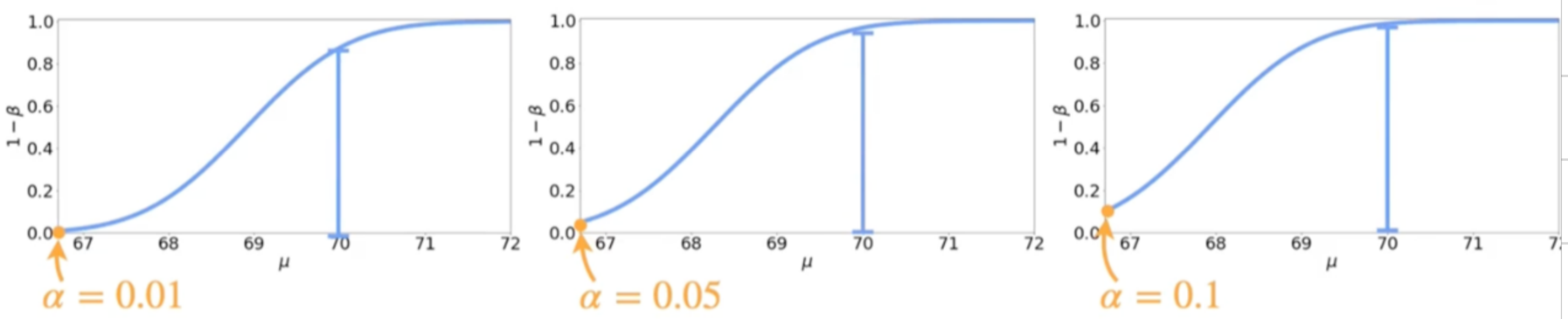
Power of the Test



Power of the Test

14

$\mu = 70$



Type I error

does the power of the test for mu equals 70.

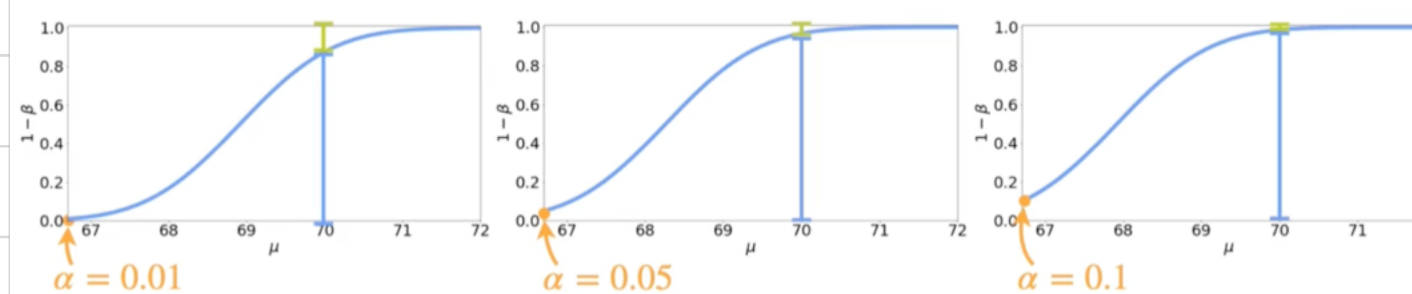
DeepLearning.AI

Power of the Test

16

$\mu = 70$

Type II error



Type I error

beta as small as you want.

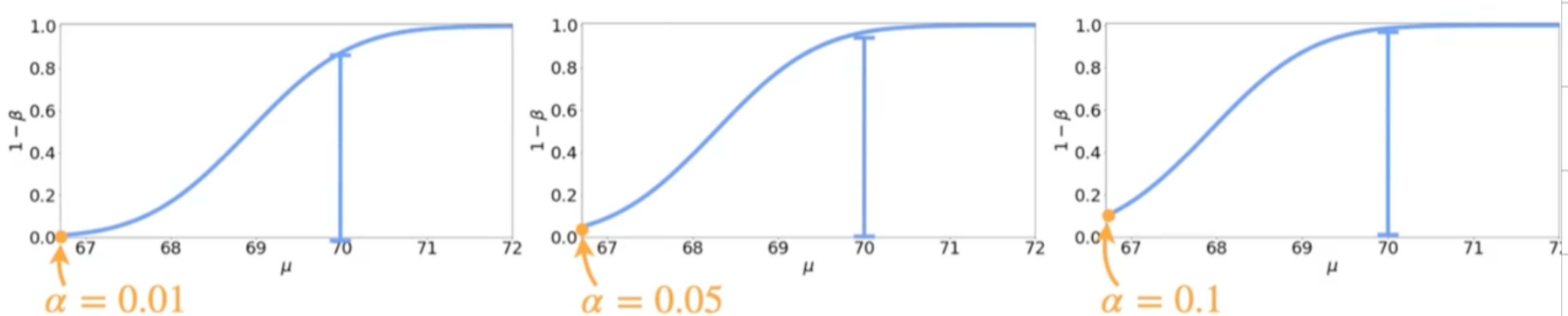
DeepLearning.AI

Power of the Test

18

$\mu = 70$

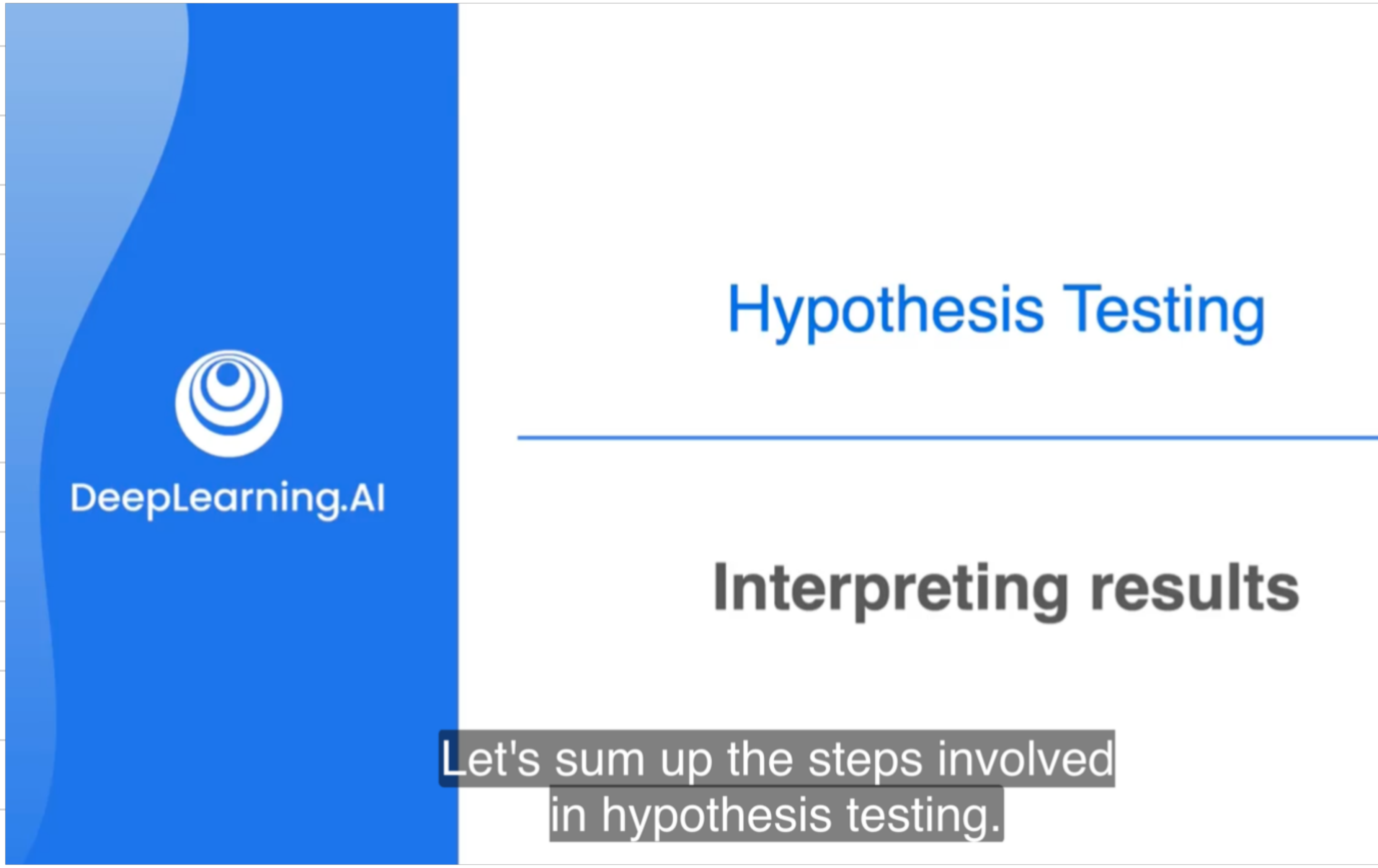
Power



Type I error

This is true for every value in the curve.

DeepLearning.AI



Hypothesis Testing: Steps and Key Concepts

1 Steps in Hypothesis Testing:

→ State Hypotheses:

→ Null Hypothesis (“H_0”): Baseline assumption (e.g., population mean height is 66.7).

→ Alternative Hypothesis (“H_1”): Statement you aim to prove (e.g., population mean height is greater than 66.7).

2 Design the Test:

→ Choose the test statistic (e.g., sample mean).

→ Set the significance level (α), typically 0.05, which is the maximum allowable probability of a Type 1 error (rejecting H_0 when it’s true).

3 Compute the Observed Statistic:

→ Calculate the test statistic from your sample (e.g., observed mean = 68.442).

4 Decision-Making:

→ Compare the p-value to the significance level:

→ If p-value $< \alpha$, reject H_0 and accept H_1.

→ If p-value $\geq \alpha$, fail to reject H_0.

Key Concepts:

→ Type 1 Error: Rejecting H_0 when it's true.

→ Type 2 Error: Failing to reject H_0 when H_1 is true.

→ Lowering α reduces Type 1 errors but may increase Type 2 errors.

Misconceptions:

→ P-value Misinterpretation:

→ A small p-value doesn't mean H_0 is false, only that the observed data is unlikely under H_0.

→ Non-Rejection Misconception:

→ Failing to reject H_0 doesn't prove it's true; it just means there's insufficient evidence to support H_1.

Null Hypothesis – No change

p-value $> \alpha$ → Reject H_1.

Alternative Hypothesis → Yes there is change

p-value $< \alpha$ → Failed to Reject H_1

Steps for Performing Hypothesis Testing

1

1. **State your hypotheses.**
   - Null hypothesis: the baseline $\longrightarrow H_0 : \mu = 66.7$
   - Alternative hypothesis: the statement you want to prove $\longrightarrow H_1 : \mu > 66.7$
2. **Design your test**
   - Decide the test statistic to work with $\longrightarrow \overline{X}$
   - Decide the significance level $\longrightarrow \alpha = 0.05$
3. **Compute the observed statistic** (based on your sample) $\longrightarrow \bar{x} = 68.442$
4. **Reach a conclusion:**
   - If the p-value is less than the significance level reject H_0

$\longrightarrow \mathbf{P}\left(\overline{X} > 68.442 \,|\, \mu = 66.7\right) \overset{?}{>} 0.05$

and accept the alternative one.

DeepLearning.AI

Important Remarks - Interpreting Tests

2

- *p*-values:
  - If $\mathbf{P}(\text{Reject } H_0 \,|\, H_0) < \alpha \longrightarrow$ Reject H_0 and then accept H_1
  - ~~The p-value represents the probability of H_0 being true.~~

While it is true that

DeepLearning.AI

Important Remarks - Interpreting Tests

3

- *p*-values:
  - If $\mathbf{P}(\text{Reject } H_0 \,|\, H_0) < \alpha \longrightarrow$ Reject H_0 and then accept H_1
  - A small p-value indicates that the probability of seeing the observed data by chance is small
- Test conclusions
  - Reject $H_0 \rightarrow H_1$ true
  - Do not reject $H_0 \rightarrow$ ~~H_0 true~~

You can only say that there is not enough evidence

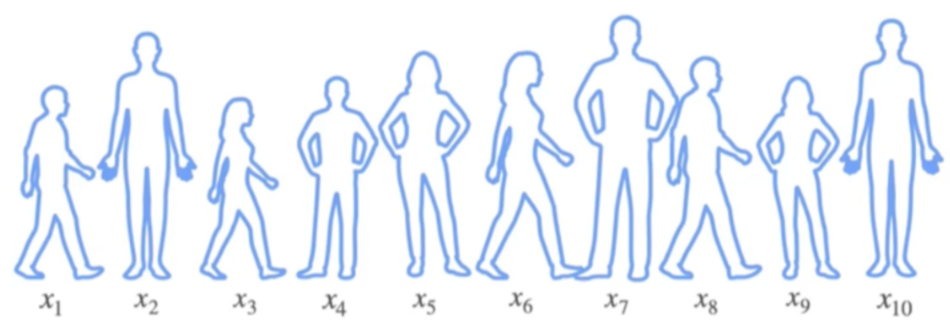
the most you can guarantee
is that there was not

DeepLearning.AI

coursera.org - To exit full screen, press esc

t-Distribution: Motivation

4



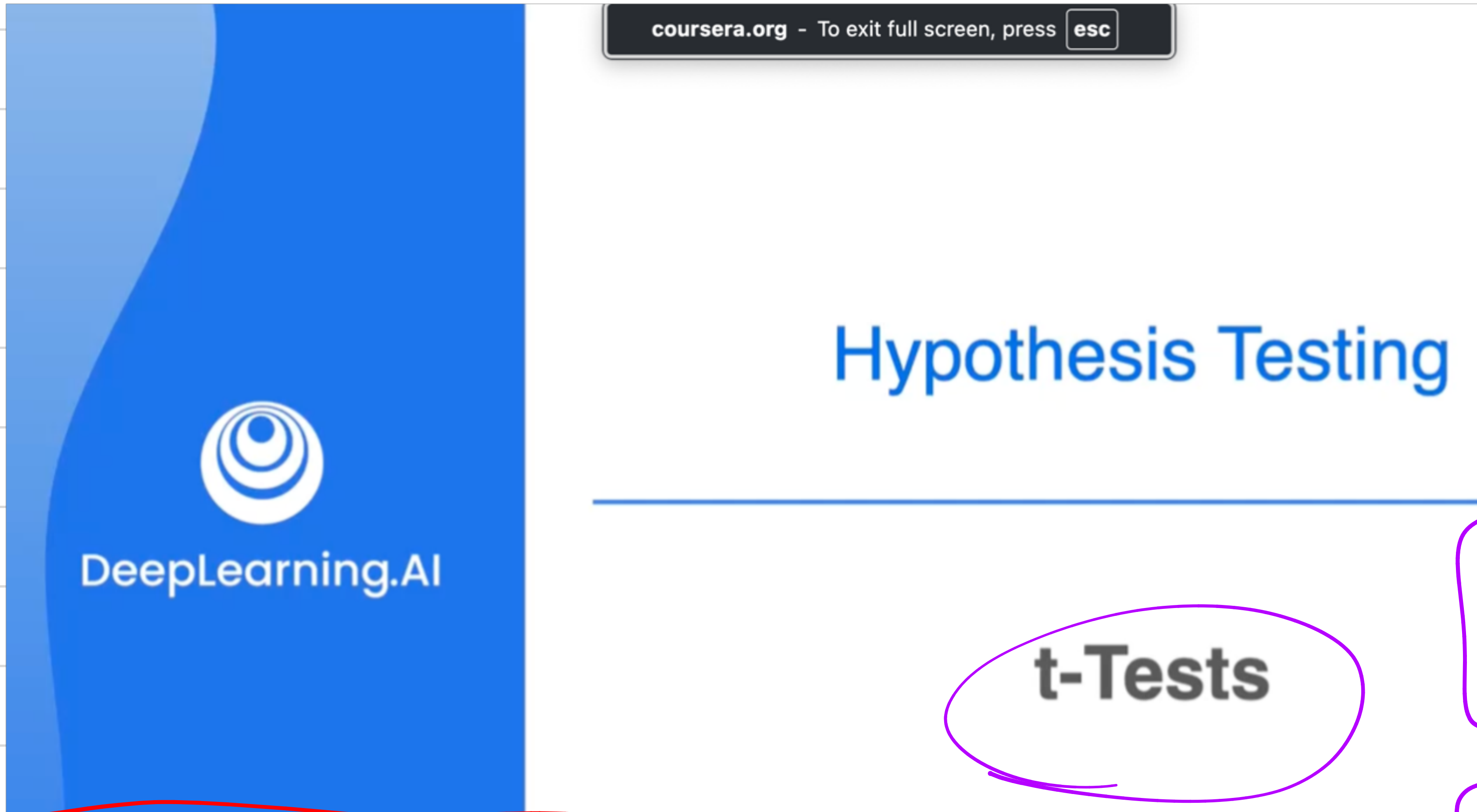
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$$X_i \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{10}\right)$$

divided by square root of 10.

DeepLearning.AI



WHY AND WHEN TO USE T-TEST

The t-test is used for hypothesis testing when:

1. Population Standard Deviation (σ) is Unknown:
   - The t-test is ideal when you don't know the population standard deviation, which is typically the case in most real-world scenarios. Instead, the sample standard deviation is used as an estimate.

2. Small Sample Sizes (n < 30):
   - For small sample sizes (usually n < 30), the sample mean might not follow a normal distribution. The t-distribution accounts for this uncertainty, as it has heavier tails than the normal distribution, making it more robust for small sample sizes.

3. When You Need to Estimate the Population Mean:
   - The t-test is used to compare the sample mean with the population mean (or another sample mean) when the population standard deviation is unknown. It helps test if there is a significant difference between the sample mean and the hypothesized population mean.

4. Handling Sample Variability:
   - The t-test incorporates the sample variance (instead of the population variance), which accounts for the increased variability that comes from using sample data. This leads to more conservative results, which is especially important when you have smaller sample sizes.

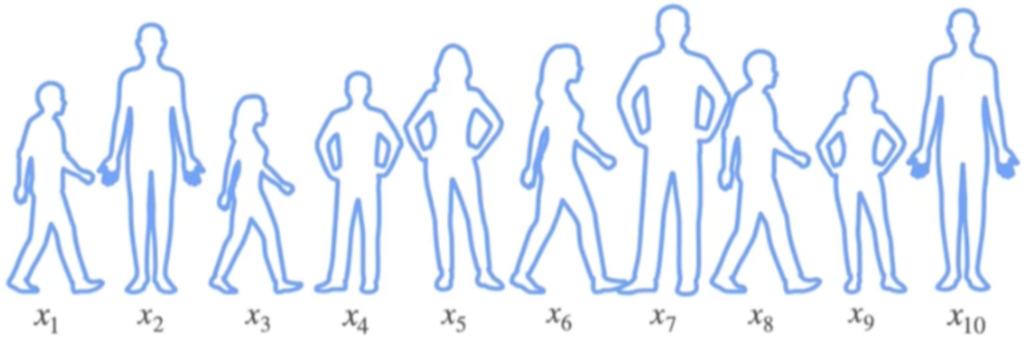
WHEN TO USE T-TEST:

- Right-Tailed Test: When you are testing if the sample mean is greater than the population mean.
- Left-Tailed Test: When you are testing if the sample mean is less than the population mean.
- Two-Tailed Test: When you are testing if the sample mean is different from the population mean in either direction (greater or smaller).

In summary, use a t-test when you have a small sample size, the population standard deviation is unknown, and you need to estimate the population mean. The t-test is a robust method that adjusts for sample variability, making it suitable for hypothesis testing in real-world data analysis.

t-Distribution: Motivation

1



66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$$X_i \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{10}\right)$$

This is fine if you know μ and σ

What if σ is unknown?

know the values of Mu and

DeepLearning.AI

t-distribution used when σ is unknown

→ t-distribution used when small-sample size used.

t-Distribution: Motivation

2

$$X_i \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{10}\right)$$

If μ, σ are known

$$\boxed{\frac{\overline{X} - \mu}{\sigma/\sqrt{10}}} \sim \mathcal{N}(0, 1^2) \text{(Standardization)}$$

Z statistic

What if σ is unknown?

T statistic

Replace σ with its estimate

$$\boxed{\frac{\overline{X} - \mu}{S/\sqrt{10}}}$$

except you divide it by
n-1 instead of n. It was

DeepLearning.AI

coursera.org - To exit full screen, press esc

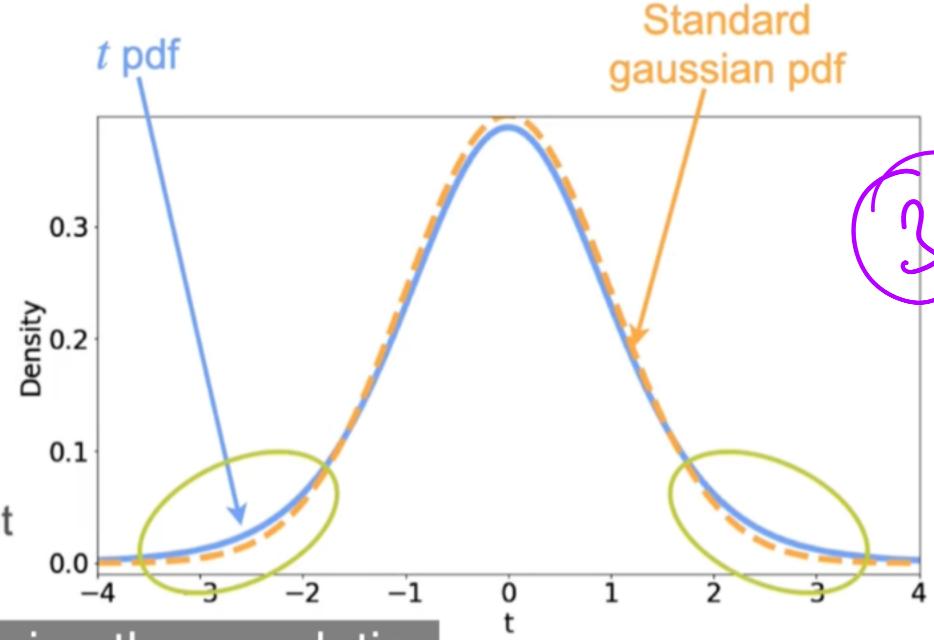
t-Distribution

$$\frac{\bar{X}-\mu}{S/\sqrt{10}}$$ follows a t distribution

What does it look like?

Still bell-shaped

It has heavier tails that account for the uncertainty introduced with the std estimation



replacing the population standard deviation

③

degree of freedom controls how heavy the tails are!

DeepLearning.AI

n-1 = dof, we do because of sample mean of an estimate population mean to correct the bias in estimating the population variance.

t-Distribution

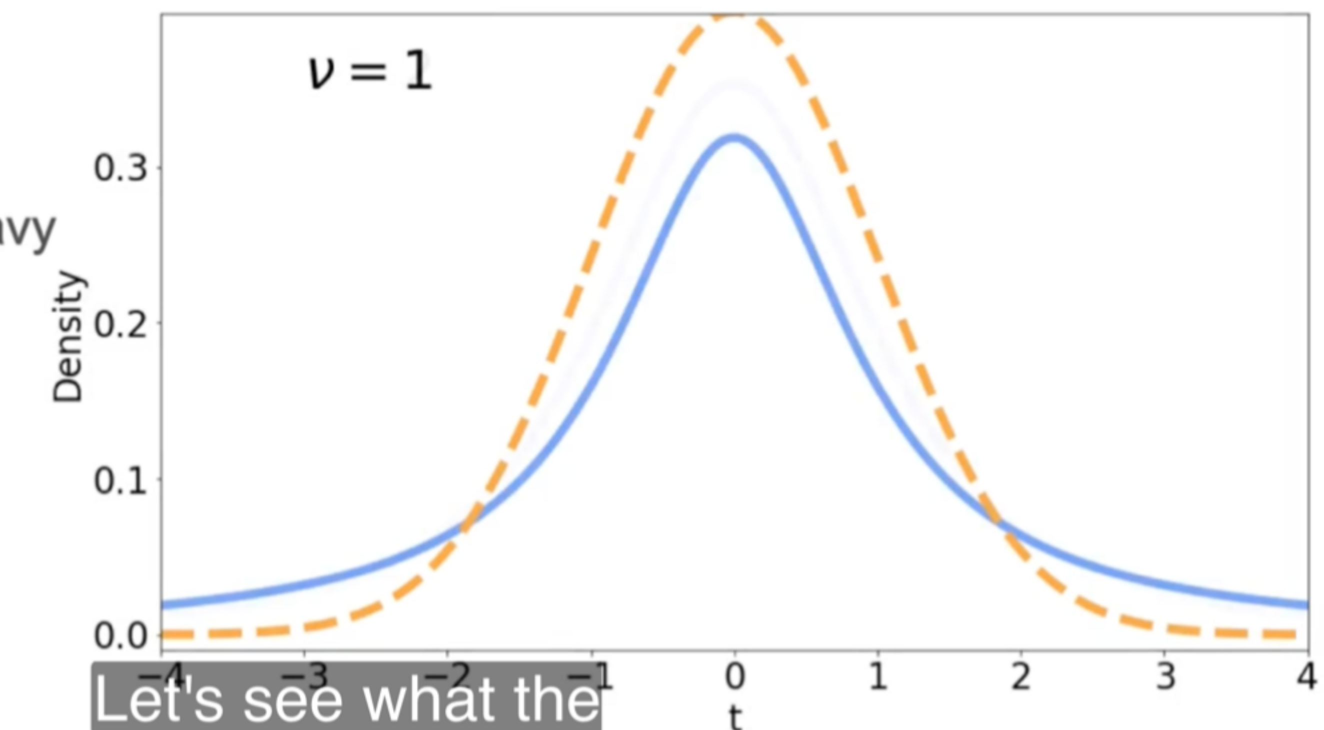
④

Parameters:

- Degrees of freedom (ν)

→ Controls how heavy the tails are

$X \sim t_\nu$



Let's see what the distribution looks like

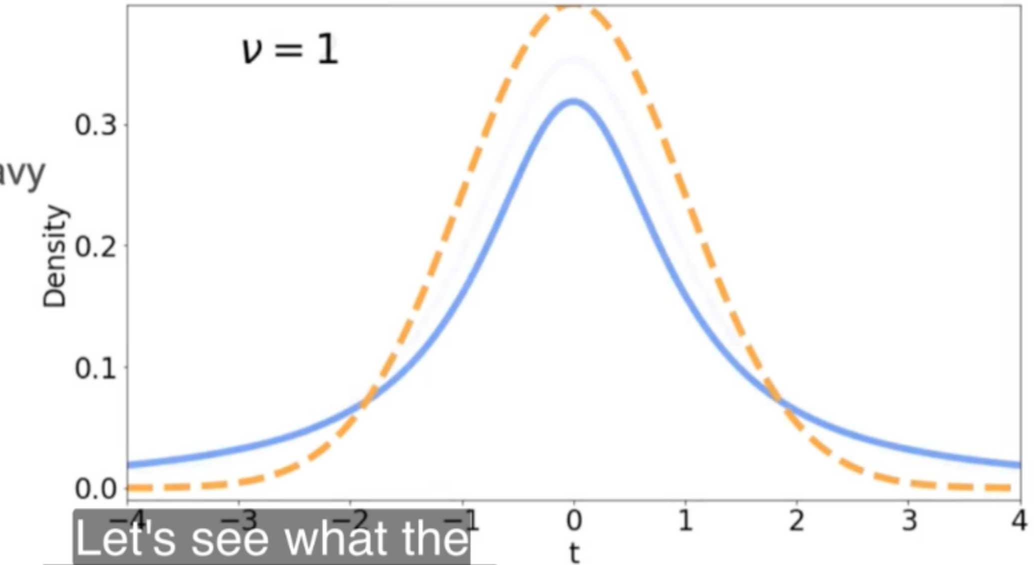
DeepLearning.AI

6

t-Distribution

Parameters:
- Degrees of freedom (ν)
  - Controls how heavy the tails are

$X \sim t_\nu$



Let's see what the distribution looks like

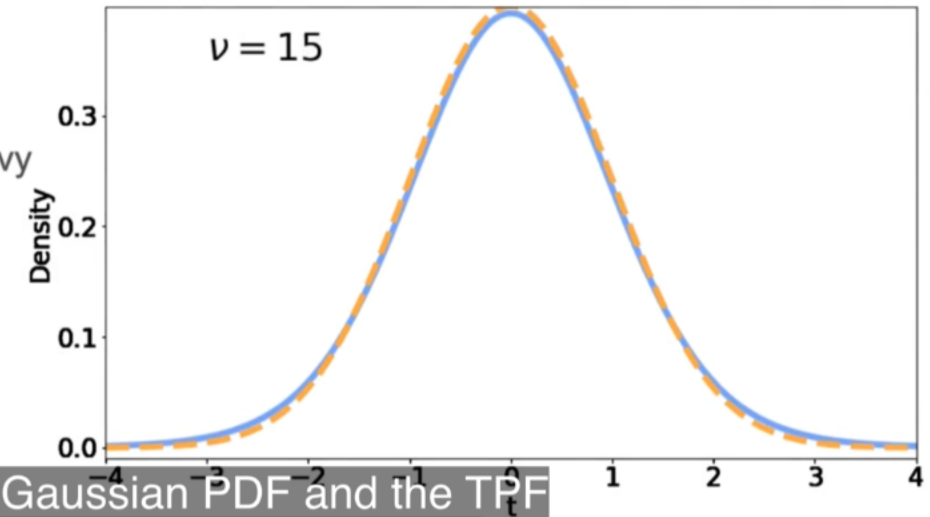
DeepLearning.AI

6

t-Distribution

Parameters:
- Degrees of freedom (ν)
  - Controls how heavy the tails are

$X \sim t_\nu$



the Gaussian PDF and the TPF look almost exactly alike.

DeepLearning.AI

7

t-Distribution and T-Statistic

$n = 10$

$\nu = 10 - 1$

$$T = \frac{\overline{X} - \mu}{S/\sqrt{10}} \sim t_9$$

Degrees of freedom (ν) = sample size -1 = $(n-1)$

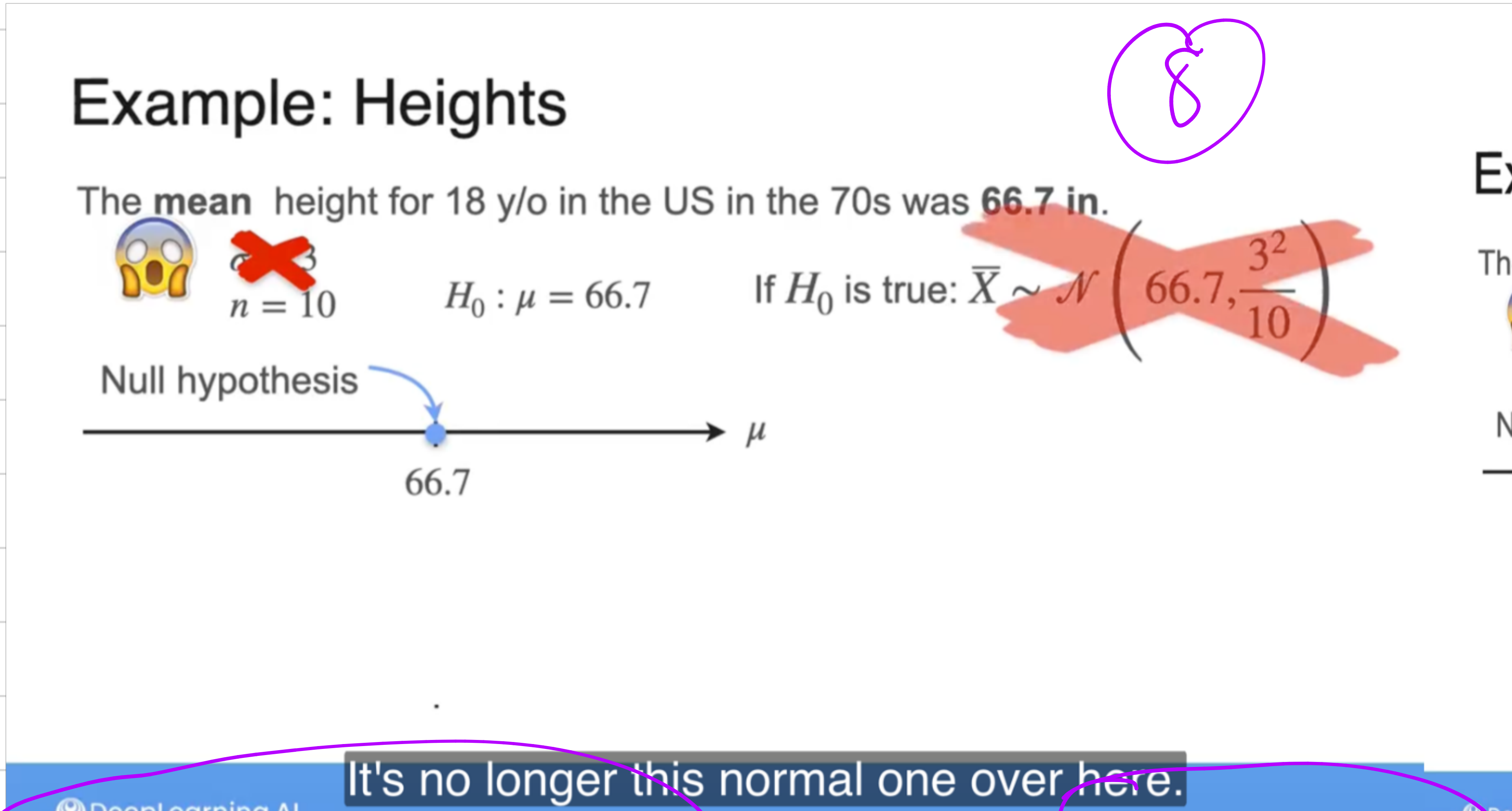
As n increases, this looks more like a $\mathcal{N}(0,1^2)$

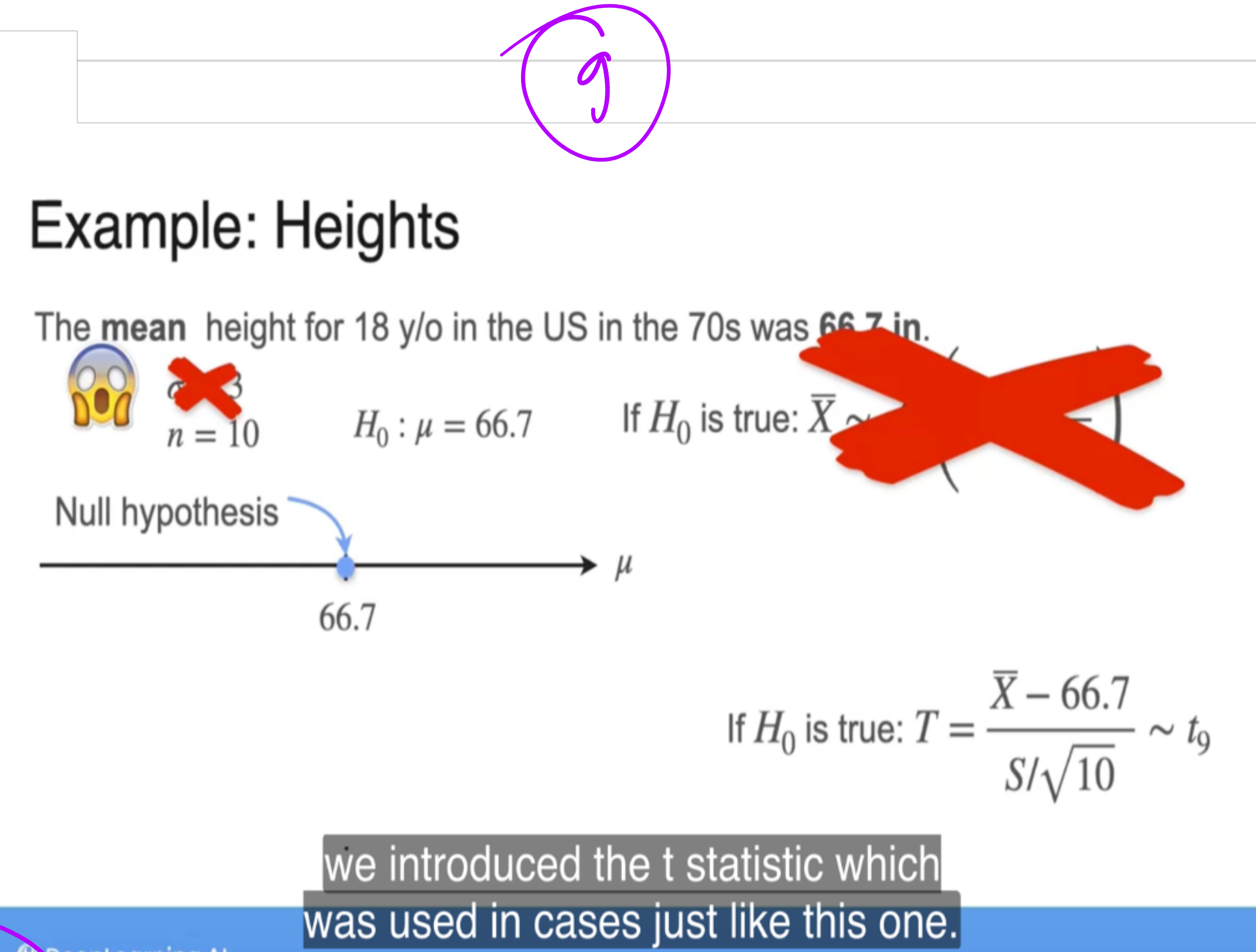
T-statistic is used when
- The population has a Gaussian distribution
- But you don't know the variance

with unknown Sigma.

DeepLearning.AI

When we don't know variance.





Hypothesis Testing with t-Statistic:

1. Right-Tailed Test:
   - Null Hypothesis (H_0): $\mu = 66.7$
   - Sample Mean: 68.442
   - Sample Size: n = 10
   - Sample Variance: 3.113
   - Calculated t-Statistic: 1.771
   - p-value = 0.0552
   - Decision: Do not reject H_0 (p-value > 0.05)

2. Two-Tailed Test:
   - p-value = 0.1103 (double the right-tailed p-value)
   - Decision: Do not reject H_0 (p-value > 0.05)

3. Left-Tailed Test:
   - New Sample Mean: 64.252
   - Calculated t-Statistic: -2.487
   - p-value = 0.0173
   - Decision: Reject H_0 (p-value < 0.05)

Summary of Results:
- Right-Tailed Test: Do not reject H_0 (p-value = 0.0552)
- Two-Tailed Test: Do not reject H_0 (p-value = 0.1103)
- Left-Tailed Test: Reject H_0 (p-value = 0.0173)

Conclusion:
Using the t-statistic with the sample standard deviation introduces more uncertainty, affecting p-values and the decision to reject H_0.

Test for proportions

1

In the videos, you learnt how to perform hypothesis testing for the mean of a Gaussian population. Another very useful example is testing for a population proportion p.

An example

Imagine that you have a coin, but you don't know whether it's fair or not. The proportion you are interested in is $p = \mathbf{P}(H)$. A possible set of hypothesis for this problem is

$$H_0 : p = 0.5 \text{ vs. } H_1 : p \neq 0.5$$

Imagine you toss the coin 20 times, of which 7 turned out heads. Your random sample consists in one random variable X ="number of heads in 20 coin flips", which has a $Binomial(20, p)$ distribution. A good estimation for the proportion is the relative frequency of heads:

$$\hat{p} = \frac{X}{20}$$

Remember that under certain conditions, the Central Limit Theorem states that $\hat{p} \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{20}}\right)$, or equivalently

$$Z = \frac{\frac{X}{20} - p}{\sqrt{p(1-p)}}\sqrt{20} \sim \mathcal{N}(0,1)$$

Z **will be your test statistic.** If H_0 is true $(p = 0.5)$, then your test statistic becomes

$$Z = \frac{\frac{X}{20} - 0.5}{\sqrt{0.5(1-0.5)}}\sqrt{20} = \frac{\frac{X}{20} - 0.5}{0.5}\sqrt{20} \sim \mathcal{N}(0,1)$$

Consider a significance level $\alpha = 0.05$. Then to make a decision you need to get the p-value for your observed statistic. With the observed sample $x = 7$, the observed statistic is

2

Z **will be your test statistic.** If H_0 is true $(p = 0.5)$, then your test statistic becomes

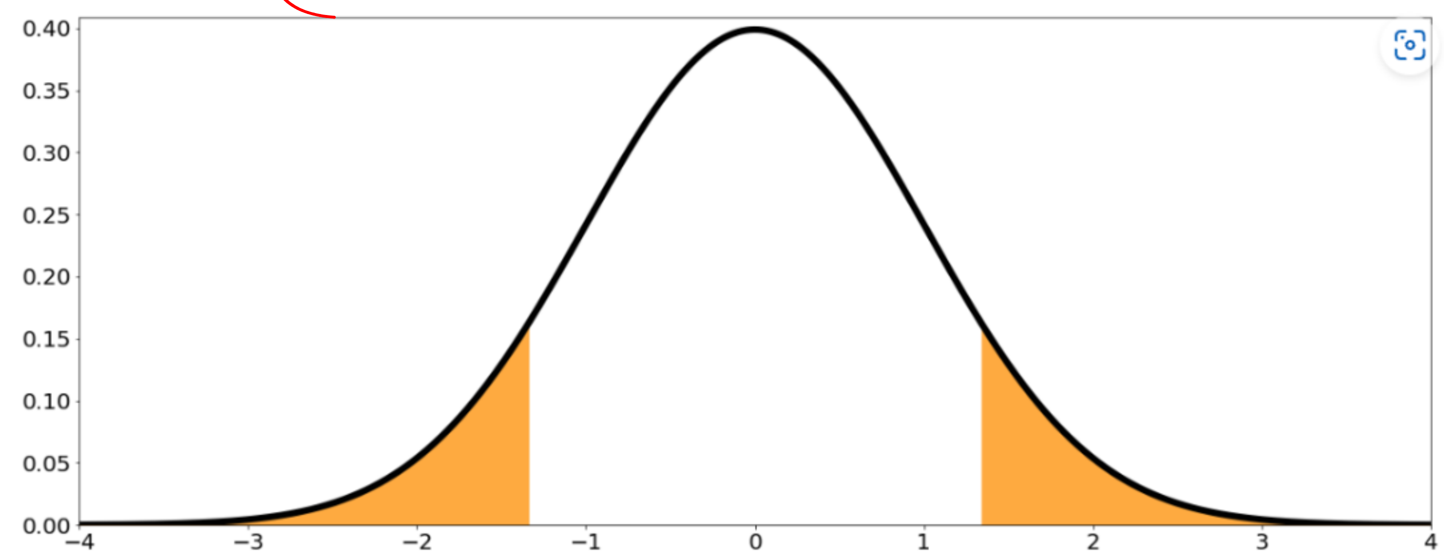
$$Z = \frac{\frac{X}{20} - 0.5}{\sqrt{0.5(1-0.5)}}\sqrt{20} = \frac{\frac{X}{20} - 0.5}{0.5}\sqrt{20} \sim \mathcal{N}(0,1)$$

Consider a significance level $\alpha = 0.05$. Then to make a decision you need to get the p-value for your observed statistic. With the observed sample $x = 7$, the observed statistic is

$$Z = \frac{\frac{7}{20} - 0.5}{0.5}\sqrt{20} = -1.3416$$

The p-value is then the probability that $Z > |z|$ or $X < -|z|$:

$$p\text{-value} = \mathbf{P}(|Z| > |z|) = \mathbf{P}(|Z| > 1.3416) = 0.1797$$



Conclusion: since the p-value is bigger than the significance level of 0.05, you do not have enough evidence to reject the null hypothesis that $p = 0.5$.

$p > 0.05 \Rightarrow$ fail to reject H_0

General case:

- p is the population proportion of individuals in a particular category (i.e. probability of the coin landing heads)

General case:

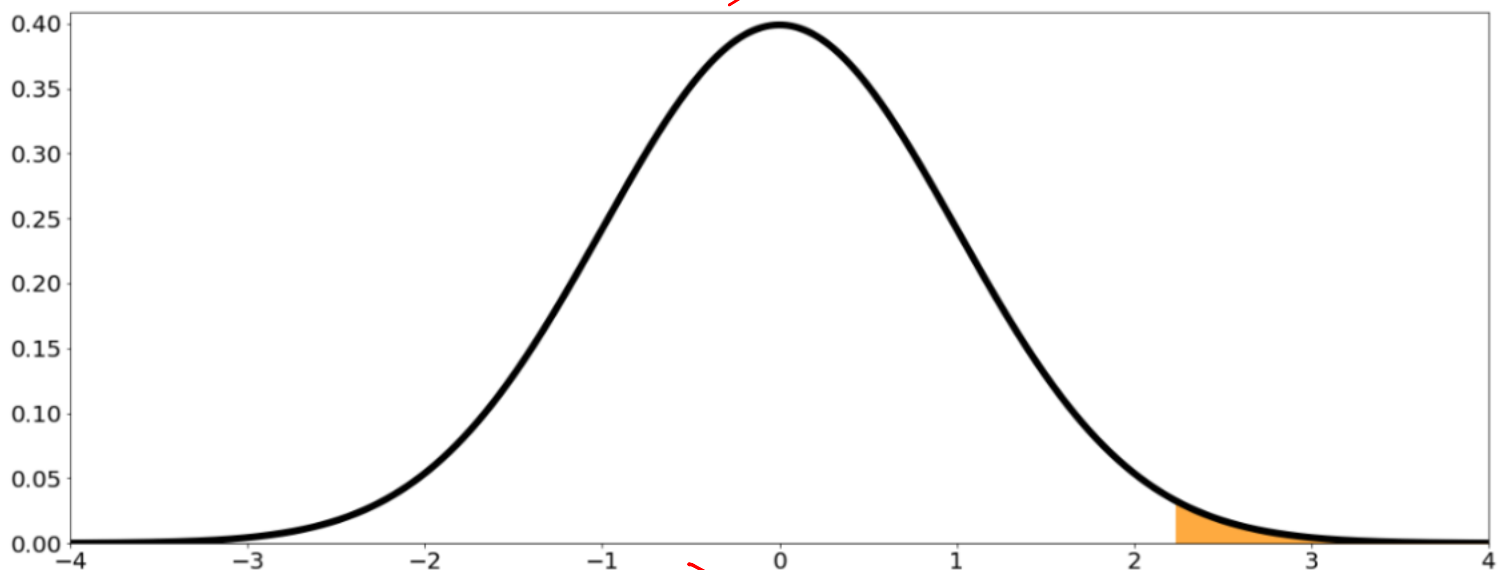
- p is the population proportion of individuals in a particular category (i.e. probability of the coin landing heads)
- p_0 is the population proportion under the null hypothesis (i.e. $p_0 = 0.5$)
- x is the observed number of individuals in the sample from the specified category (i.e. number of heads)
- n is the sample size (i.e. number of coin toss)
- $\hat{p} = \frac{x}{n}$ is the sample proportion for the observed sample x.

Then, $Z = \frac{\frac{X}{n} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \sim \mathcal{N}(0,1)$ is the test statistic for comparing proportions, and $z = \frac{\frac{x}{n} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ is the observed statistic.

Depending on the type of hypothesis, you have different expressions for the p-value:

- Right-tailed test: $H_0 : p = p_0$ vs. $H_1 : p > p_0$:

  $$p\text{-value} = \mathbf{P}(Z > z)$$



- Left-tailed test: $H_0 : p = p_0$ vs. $H_1 : p < p_0$

  $$p\text{-value} = \mathbf{P}(Z < z)$$

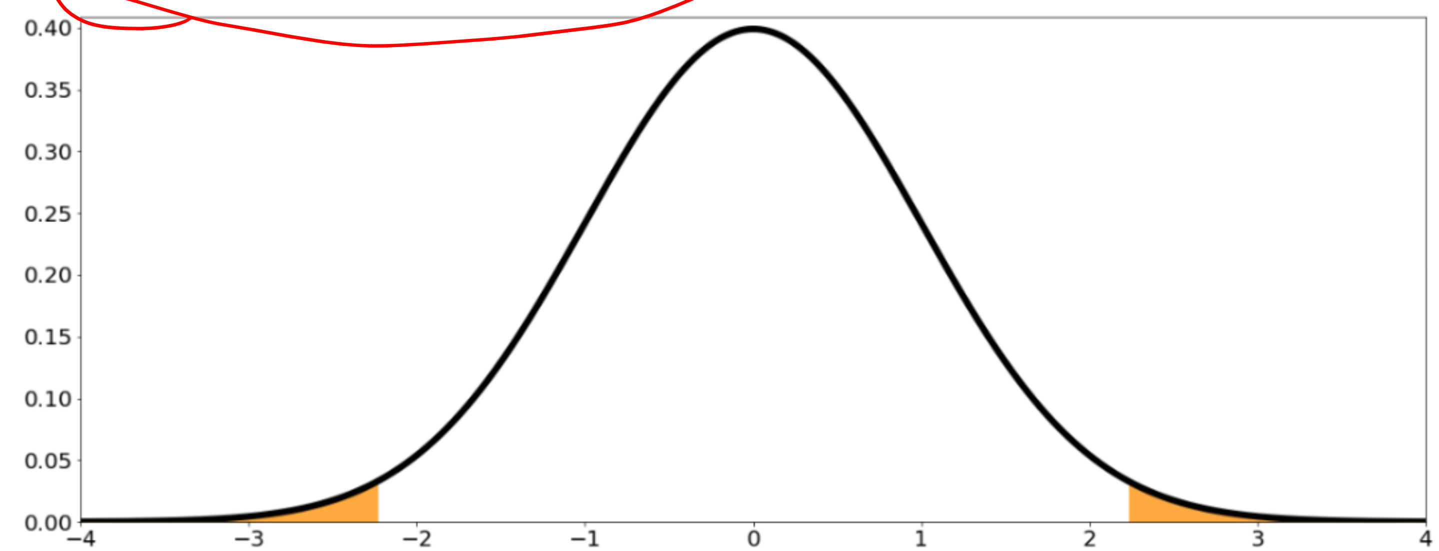
- Left-tailed test: $H_0 : p = p_0$ vs. $H_1 : p < p_0$

  $$p\text{-value} = \mathbf{P}(Z < z)$$

- Two-tailed test: $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$

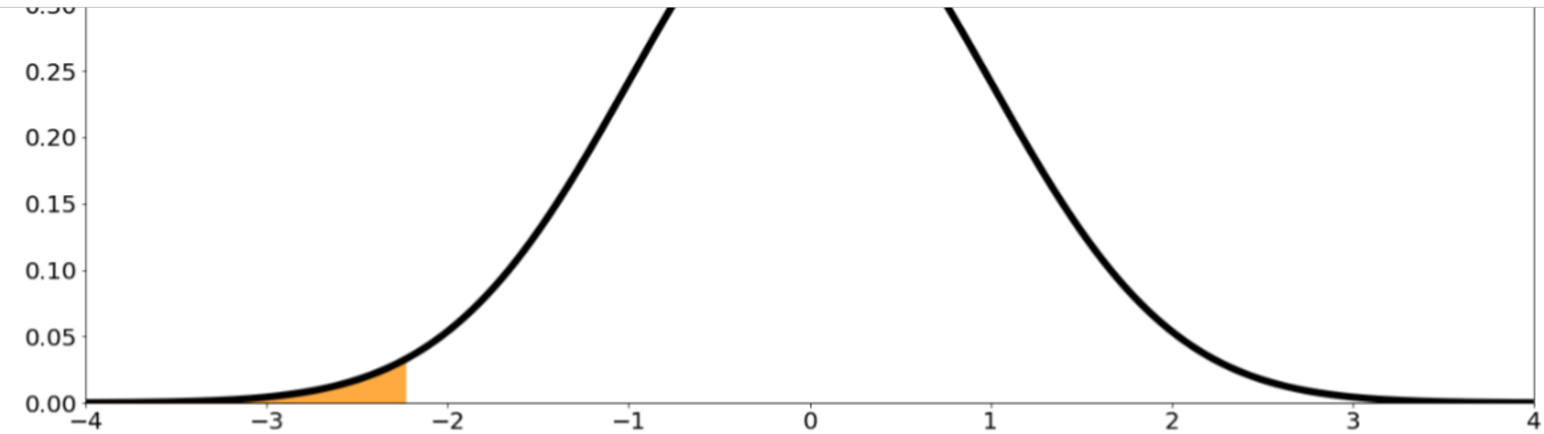
  $$p\text{-value} = \mathbf{P}(|Z| > |z|)$$

$p = p_0$ vs $H_1 : p \neq p_0$



For this results to be valid, the following conditions need to be satisfied:

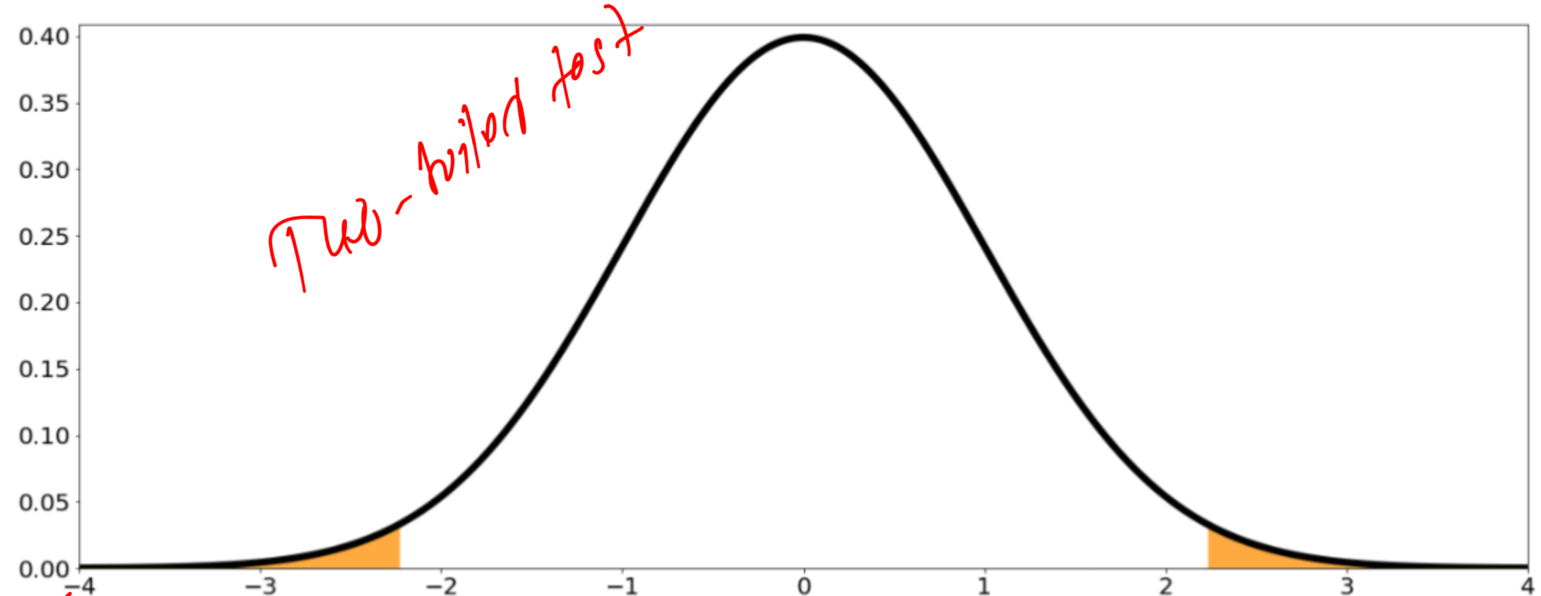
- The population size needs to be at least 20 times bigger than the sample size. This is necessary to ensure that



- Two-tailed test: $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$

  p-value $= \mathbf{P}(|Z| > |z|)$

Two-tailed test



For this results to be valid, the following conditions need to be satisfied:

- The population size needs to be at least 20 times bigger than the sample size. This is necessary to ensure that all samples are independent. This condition is not needed in situation like the coin toss, where independence is inherent to the experiment.
- The individuals in the population can be divided into two categories: whether they belong to the specified category or they don't
- The values $np_0 > 10$ and $n(1 - p_0) > 10$. This condition needs to be verified so that the Gaussian approximation holds when the assumption that H_0 is true.

Sources:

→ Probability and Statistics for Machine Learning and Data Science.